CAPCO

JOURNAL

THE CAPCO INSTITUTE JOURNAL OF FINANCIAL TRANSFORMATION

DIGITIZATION

#47
04.2018

JOURNAL

THE CAPCO INSTITUTE JOURNAL OF FINANCIAL TRANSFORMATION

RECIPIENT OF THE APEX AWARD FOR PUBLICATION EXCELLENCE

CONTENTS

ORGANIZATION

CURRENCY

SECURITY

ORGANIZATION

PREVIOUS EDITIONS OF THE CAPCO JOURNAL OF FINANCIAL TRANSFORMATION ARE AVAILABLE AT **WWW.CAPCO.COM/INSTITUTE**

- Implications of robotics and AI on organizational design
- The car as a point of sale and the role of automotive banks in future mobility
- Fintech and the banking bandwagon
- Can blockchain make trade finance more inclusive?
- The aftermath of money market fund reform
- Costs and benefits of building faster payment systems: The U.K. experience
- Household deformation trumps demand management policy in the 21st century

Implications of robotics and AI on organizational design

PATRICK HUNGER | CEO, Saxo Bank (Schweiz) AG
RUDOLF BERGSTRÖM | Principal Consultant, Capco
GILLES ERMONT | Managing Principal, Capco

ABSTRACT

Although robotics and artificial intelligence (R/AI) create opportunities to improve operational efficiency within organizations, they are also seen as threats to jobs. The idea that machines can now do what white collar workers have been doing for decades has raised concern, and many are questioning the ability of humans to compete with computers. In this article, we will explain that these new technologies are not only intended to cut costs and headcount in repetitive tasks, but also enable businesses to become even more innovative by refocusing on the strengths of their human workforce.

1. INTRODUCTION

In this article, we provide a different perspective on robotics and artificial intelligence (R/AI) and investigate the implications that these new disciplines will have on organizational design. Our intention is not to undertake an academic analysis of the issues, but to discuss and share our experiences and thoughts in a practical manner with those who are facing these challenges within the business world. To this end, we interviewed several executives at Saxo Bank to gain their perspectives on how robotics is used, its implications on the organization so far, and what they foresee could happen in the future. While we do not claim to have all the answers to the complex questions that R/AI raises, we hope that this article will trigger more forward-looking reflections on their long-term implications.

2. DRIVERS AND CHALLENGES FOR R/AI DEPLOYMENT

An important question that needs to be addressed is: Why are businesses investing in robotics? The financial services (FS) industry has particularly shown an interest. Banks and insurers were among the first businesses to launch large scale robotic process automation (RPA) projects. But why?

The FS industry has, it seems, found in robotics a promising way to further automate activities that were previously only performed, and possible to perform, by humans. While the immediate value proposition focused on operational efficiency, cutting down headcount as manual work is transferred to machines, businesses are now realizing that there is more to be achieved from RPA, be it in supporting scalability or enabling new value adding activities.

Numerous banks have built back office functions using off/nearshore resources and have overlooked how inefficient they were simply by focusing on lower labor costs and relying on human ingenuity to work around complexities in their application and integration architecture.

The recent wave of RPA roll-outs across the industry has been driven by tactical cost efficiency targets, with a strong focus on automating rule-based back office activities. In that context, RPA has been viewed as an alternative to complex IT integration projects and other near/off-shoring strategies.

After all, why spend time and resources re-engineering processes and underlying systems if you can quickly and cheaply fill the gaps with RPA? As once stated by Bill Gates: "The first rule of any technology used in a business is that automation applied to an efficient operation will magnify the efficiency. The second is that automation applied to an inefficient operation will magnify the inefficiency."

The idea that banks could easily replace hundreds, if not thousands, of human workers, or fix a fragmented IT architecture using robots quickly clashed with reality: lack of process standardization, misalignment between business and IT teams, and ever-changing application landscapes are some of the roadblocks that robots have found on their way to operational domination.[1]

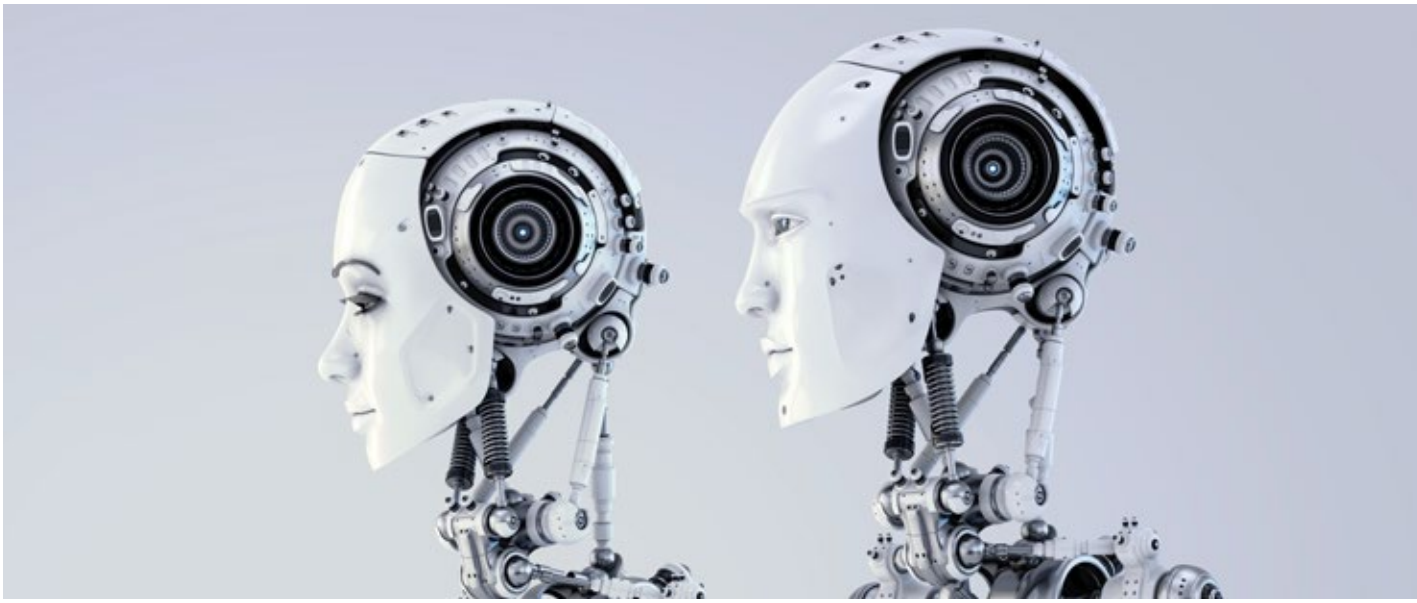
As companies went through the effort of configuring and launching robots, they started to realize that a successful RPA implementation requires a fresh look at how processes are designed, how teams are organized, and even why they have built certain capabilities in the first place.

When it comes to AI, the picture is slightly different. Unlike RPA, the business case for AI is not self-explanatory. RPA is very intuitive to understand and easy to turn into metrics, i.e., I will replace a person performing X tasks per day at the cost of Y per year by a machine performing more of the same tasks at a lower cost. In more ways than one, the case for RPA is process-centric. Not so much for AI.

AI has come a long way since the initial concept was formalized in the first half of the twentieth century. However, we have yet to see a fully functioning general purpose of AI, one that is able to learn to do everything a human does. What we do have, though, are specialized AI and machine learning systems being applied opportunistically to create point solutions.

These are becoming key contributors to decision-making processes, performing analysis that no one had time to perform previously. The business value of such systems lies more in how they help people achieve desired business outcomes than reducing headcount.

[1] The Volume 46 edition of the Journal of Financial Transformation has a number of articles discussing the challenges many organizations face in applying RPA.



Christian Busk Hededal, Saxo's Head of Big Data & AI, states: "With AI and ML, we have three main areas of focus: the AI-based information engine, predictive analysis, and personalized marketing. We want to be able to deliver the right information to the right clients as part of our client service."

The implications of robotics on organizations are already visible, even for companies that have only just started their robotization journey.

As teams are downsized and human points of contact are removed from the process execution chain, one can wonder how team-to-team communication and roles, such as "team lead" or "team manager," will be impacted once large portions of the work is handed over to robots.

In parallel, new roles are emerging, such as developing and monitoring the robots, ensuring synchronization with IT teams, periodically reviewing the robots' output against business expectations, and continuously assessing potential to automate further.

In the short term, accountability of the process managers (PMs) will become even more important. Not only will they be in charge of defining efficient processes, they will also need to ensure that robots are performing the right tasks. PMs will gather and maintain knowledge from across the organization and work with the technology teams to keep tweaking the robots and maximize their utilization.

In the long run, this also creates expectations on people to reallocate their time to more value-adding activities; in a sense moving away from rule-based work to more outcome-based contributions to the business. This is where the challenge lies: how do we find value-adding activities when we have been obsessed with cost cutting?

3. EVOLVING FROM RULE-BASED TO OUTCOME-FOCUSED ORGANIZATIONS

What if everything we do more than once could be handed over to machines?

With the commoditization of robotics solutions, every business will very soon be able to automate most, if not all, of their rule-based tasks. As such, efficient execution alone will not be a competitive advantage anymore. In turn, remaining ahead of the competition will require organizations to focus on desired business outcomes rather than measuring process execution and adherence. This will demand that people pivot to a more outcome-based mindset, using probabilistic tools based on non-absolute truth as opposed to the prevalent rule-based deterministic way of solving problems.

While in the short-to medium-term robotics automation can provide a competitive advantage for businesses that are able to apply it to raise their operational efficiency above their peers, there is little to no doubt that, like other technologies before, it will be commoditized over

time. It will not be too different to the history of electricity. In the early years of the industrial era, factories had to build their own dedicated power plant, however, as electricity production and distribution grew more stable and efficient, it became cheaper and more sensible to just plug into the general grid.

If we try and look to the future, we can easily envision a world where automation is widely available; and we are only referring here to a few years, not decades. Cloud-based robot farms could be accessed on-demand to execute rule-based activities without the need for businesses to spend time and resources building their own internal capabilities.

When the technologies required to remove the human workforce from rule-based activities become widely available, efficient execution will be a commodity and not a competitive advantage anymore.

This model would shift the baseline for competitive advantage. If everything that is rule-based and performed more than once is automatable via robotics, and if the technology to do so is available to all, then businesses must differentiate on something else. Being a bit provocative, efficient execution could be looked upon as an "old world problem," something that has been solved and can now be taken for granted, the same way that businesses take electricity for granted and do not feel the need to build their own power plants anymore.

In this scenario, businesses will have to compete on new grounds, invest into what we will qualify as "new world solutions."

And, this is where the real challenge lies for most organizations. Decades of focusing on using rule-based approaches to solve business questions has created biases that are hard to overcome.

Using rule-based deterministic processes (i.e., a set of "if/then" statements using predefined triggers and resulting in a predefined outcome) is comfortable because they are easy to follow, easy to measure and control, eliminate surprises, and are not people-dependent. That is why such processes are the first ones to be outsourced or moved to near/offshore centers and, now, robotized. Reconciliations is a good example of such a rule-based process, starting by identifying differences between bank statements and client records and performing the needed adjustments based on predefined criteria.

This in turn fosters a mindset that focuses on improving existing codified processes that lend themselves quite well to further rule-based optimization rather than exploring new questions that are not so easily solved through a list of "if/then" statements.

Rule-based processes, designed to reach a predefined outcome in a repeatable and controllable manner, put much focus on defining the steps and control points. As routine sets in, people can lose sight of why the process was designed in the first place to focus on repetitive execution and eventually mistake process execution for value creation. One can always fall back on being compliant to justify how one has created value.

> "By freeing us from mundane tasks, robotics allow operational staff to focus on adding value to the company."
>
> Gerard Lelliott, Saxo's Global Head of Operations

Gerard Lelliott, Saxo's Global Head of Operations, suggests that "By freeing us from mundane tasks, robotics allow operational staff to focus on adding value to the company by using creativity and lean skills on developing new products and designing better scalable processes, resulting in a better client experience."

With machines solving old world problems, the real competitive advantage will rely on organizations' ability to invent new world solutions: new services, offerings, and products built from a customer centric-perspective and freed from the constraints of human execution.

Gerard Lelliott goes on to say that "Most people in the back office would rather be involved in product and process development rather than doing manual processing."

Looking at your business with a deterministic mindset assumes that you have absolute truths at your disposal. These absolute truths can come in the form of proven facts, such as a change in current interest rate in a given country. But not all information we process is fact-based. A lot of it comes from assumptions, such as expected reactions to the change in interest rate. And for the most part, humans are good at making sound assumptions that turn out to be proven correct, so we tend to deal with these assumptions as if they were absolute truth. However, businesses can come across problems that

require processing an overwhelming amount of facts, or for which making assumptions is nearly impossible. When this happens, we have a natural tendency to deem the question impossible to solve and we move on to simpler ones, ones that we can work out via manageable datasets and assumptions – i.e., old world problems.

Prabhu Venkatesh, Saxo's Head of Data, says: "We need to learn to deal with non-absolute truth and the probabilistic nature of things around us."

Over the years, organizations have been trained to use rule-based approaches to solve business problems and over time it has influenced the type of questions/problems that they put their focus on; filtering out the ones that do not lend themselves very well to rule-based thinking. This is where businesses can miss out on untapped value, by focusing only on problems that are easily solved by rules and absolute truths.

With the advent of AI and machine learning, we can now use machines to support a more probabilistic approach to solving business questions and even start working on problems that we had previously elected to ignore because they would have been too complex/time-consuming to solve. A good example of a problem that does not lend itself well to rule-based analysis is predicting customer behavior. Banks are monitoring customer activities to comply with, for example, AML regulation, capturing massive amount of data that can then be used to better understand customers. So much data is available in fact, that it would be impossible for humans to crunch it into something usable for the business.

Christian Busk Hededal, Saxo's Head of Big Data & AI, says that "Often, it would be far too complex to try to understand our clients by applying a rule-based approach. In the example of fraud detection, it is often subtle correlations in seemingly unrelated data that can make the identification. Here, machine learning is superior."

As a response to that, Saxo has developed a machine learning algorithm to rank leads based on the probability to convert into actual sales, helping relationship managers optimize their time with prospective clients.

As this example shows, AI can create value for businesses by enabling them to be more data-driven, using machines to perform tasks that would have been too complex or time-consuming for humans, removing the need to manually process large data sets and make questionable assumptions. And in doing so, refocusing the human workforce on doing what it is best at: designing solutions to ever more complex questions through a mix of intuition and sound assumptions – i.e., "new world solutions."

> "With AI / ML you don't want to be behind the curve by being under-skilled or not invest appropriately. In the end, you need few but very good people to make it work."
>
> Christian Busk Hededal, Saxo's Head of Big Data & AI

4. RETAINING THE ABILITY TO EVOLVE ORGANICALLY AFTER AUTOMATION

We discussed the value of having access to ever more efficient execution capacity, but what about the ability of a business to evolve and change the way it operates?

Like every piece of technology, robotics works as instructed by humans, which leads to the question of how businesses can keep innovating and improving while relying increasingly on robots. The effort needed to build robotics solutions and the time they free up, create opportunities for the remaining human workforce to build the mindset and methodologies needed to continuously look for improvements and design new value-adding activities.

Unattended robotics automation can act as a fixed prosthetic organ or limb in the way that it is very efficient at executing predetermined tasks in a predetermined way. Over time, an organization can simply forget how the robots work and why they were deployed in the first place. Banks are already facing similar challenges with an aging Cobol developer community retiring, leaving newer IT staff without the knowledge and experience to maintain legacy systems that were built decades ago.

So, how might a company retain its ability to evolve organically while still being a heavy user of robotics solutions?

4.1 Breaking organizational silos

Many of us who work in a computerized environment have been hearing for the last 25 years that the key to sustaining organic evolution is to foster alignment between the business and the technology teams. While that sounds reasonable on paper, traditional IT systems

have been managed as monolithic blocks designed and maintained by dedicated teams. As a result, many companies have built a cultural gap between their business and IT teams, with the latter focusing on keeping the lights on for the former. Robotics demands that these teams work together as closely as possible to ensure that they not only keep up with the evolution of the business, but also foster it.

Patrick Hunger, CEO of Saxo Switzerland, says in this regard that "From an impact perspective, it is less imperative how your organizational setup is intellectually designed. What creates organizational mindshare is 'human transactions'; all guided by a collective and culturally well anchored business purpose."

Prabhu Venkatesh, Saxo's Head of Data, says: "We have a bi-directional, collaborative model, with free flow of ideas and information between tech and business teams. Tech knows what's possible, business knows what's useful – magical products are born in that intersection."

Christian Busk Hededal, Saxo's Head of Big Data & AI, suggest that "You need to have clear dialogue with IT and business as equal parts of the company. At Saxo, we have decided to have the machine learning and AI development team as an integral part of the business organization to bridge the gap. We have the mentality of being a data-driven organization with close alignment between IT and business."

Saxo has taken a very pragmatic approach to robotics automation, seeking to learn from the first movers and then carefully plan its journey. An example of that is how Saxo is combining lean methodology as part of process automation work to bridge work across organizational silos and aligning business and IT.

Gerard Lelliott, Saxo's Global Head of Operations, says that "We don't want people to think of robotics and lean as two different things, for Saxo they need to be used together to drive our scalability. This way we can build the right mindset to continuously improve." Nino Adamia, Saxo's Head of Business Process Management, adds that "Building a lean and improvement mindset is needed to make a robotics initiative efficient. And it also made people generally very positive about robotics."

4.2 Build a data-driven culture

Robotics technologies can create a data-driven culture for continuous improvement initiatives, enabling the organization to target the right pain points and measure value from improvements initiatives more accurately. RPA provides detailed, step-by-step execution data, and it is up to the organization to consume this data, creating a constant feedback loop for improvement.

Prabhu Venkatesh, Saxo's Head of Data, states that "We are using data and analytics to help operations identify bottlenecks in the customer onboarding process, bottlenecks that could hinder growth. We want to avoid piece-wise automation that fragments the work between humans and machines. Automation works best when there are few and clean human-machine interfaces."

Another effect of robotization is the ability to expose data and KPIs to a wider group of stakeholders. As simple as it sounds, being transparent with execution data creates a common understanding among different teams about what is happening in the company and how everybody's work impacts it. It helps every single employee to keep the big picture in mind, while becoming more data driven in decision-making. It also creates an outcome-based culture where humans act upon visible issues rather than relying on the process itself.

Prabhu Venkatesh goes on to say that "Data is flying at you through the air constantly, so you might as well do something with it right away, instead of storing now and analyzing later."

4.3 Set right expectations and commit to it

Roy Amara, President of the Institute for the Future, once said that "We tend to overestimate the effect of a technology in the short run and underestimate the effect in the long run"

Like many other technologies before, robotics has set sky-high expectations in terms of what it can achieve. Yet, companies need to acknowledge that it is no silver bullet and that machines and people will coexist for the foreseeable future.

Prabhu Venkatesh stresses that "The ones who have been most successful with AI are the ones who have had the right expectations."

Companies must learn to use machines for what they are good at, i.e., processing large volumes of data with little judgment. Given the amount of data that we generate every minute of every day, the portion of actual clean and verifiable data that a business can

use has decreased. Machines are perfectly suited to help humans deal with this paradox, as they can quickly reduce the noise and enable people to put more focus on their objectives. In the words of Saxo's Prabhu Venkatesh Head of Data: "Automation is allowing humans to do more of what humans are good at doing."

Christian Busk Hededal, Saxo's Head of Big Data & AI, adds: "With AI/ML you don't want to be behind the curve by being under-skilled or not invest appropriately. In the end, you need few but very good people to make it work."

An example of the above can be found in the automated bond trading system created by Saxo Bank. Saxo's Head of Fixed Income, Simon Fasdal explains: "We are relentlessly removing manual points in the value chain."

The solution created by Saxo uses RPA to replicate everything a trader does in a given market, just much faster and more reliably. Up until now, trading bonds has been mostly a manual affair in a very fragmented and non-transparent market, from the client contacting a trader to place an order, to the trader dealing with brokers, coming back with a price offer until finally the trade is settled. Now, all a client needs to do is place an order in a certain price range and the system autonomously screens brokers to identify the best match on the market with much more transparency and faster execution of orders.

The speed and efficiency of robots now allows bond trading to be almost as fluid as equity trading. As Fasdal states: "The automation of bond trading will impact and change the organizational structure by cutting out excessive touchpoints in the value chain. This will dramatically change the roles and setup of the current teams working in that area. The efficiency of the system is way above that of the manual value chain."

The benefits extend beyond cost efficiency, with reduced spreads, increased transparency and regulatory compliance, and most of all scalability.

Simon and his team are now looking into adding AI and ML components to the system to analyze failed trades and mismatch between clients' spreads and final prices. All of this is done in close collaboration with the IT and operations team to ensure that improvement and innovations can be properly scaled across the organization.

"This is the benefit of our collaborative corporate culture. We work with controlled anarchy," concludes Simon.

5. HARNESSING THE TRANSFORMATIVE POWER OF ROBOTICS AT AN ORGANIZATIONAL LEVEL

Realizing long-term benefits of robotics solutions requires businesses to properly manage the transformation of their workforce, building the internal structures to foster not only robotics adoption but also the need to constantly evolve. As such, one can envision the organization of the future as an environment where humans focus solely on change while everything that is executed more than once is left to machines.

If we pause for a moment and reflect on the topics we have discussed so far, we realize that a successful robotics program looks more like a top-down re-engineering of the organization than a traditional technology or process transformation.

From that perspective, there are key areas that businesses need to address to both navigate the challenges created by robotics and realize the long-term benefits.

Start from the top:

- Leaders need to be fluent in robotics so that they can not only create and advocate a compelling robotics vision and journey for the organization, but also articulate the strategic importance for the enterprise
- Empower robotics advocates who will become the day-to-day change agents.

Establish a robotics change engine:

- Establish a strong governance to manage the delivery of robotics solutions against expected business value and constantly investigate new ways for the business to benefit from robotics
- Challenge the status quo by overcoming organizational and process boundaries that are rendered obsolete by robotics.

Sustain organizational change:

- Support operational managers with the practical methodologies and tools needed for the daily management of a mixed workforce made of both humans and machine. In addition, help managers and staff cope with the anxiety that come with any change
- Involve HR early on to provide recommendations on redeployment of the human workforce.

These topics are nothing new and they are not specific to robotics. They are the recommendations that come with any organizational transformation. We are just stressing the need to look at robotics not only as a change in technology or processes, but a more fundamental change in the organizational design that needs to be recognized as such by the leadership across the business.

Patrick Hunger, CEO of Saxo Switzerland, states that "It is the role of the leadership to consciously design a 'transactional corporate organism' that nourishes through inclusion – and not separation – innovation and performance in a human-machine ecosystem."

Done right, robotics can drive change in all parts of the organization. There is a distinct possibility for the robotics change engine mentioned above to become the actual business of the future: cross-functional teams constantly investigating new ways of creating value for the customers and the company, while leaving the execution to robots.

As we left the trading floor at Saxo, we passed by a team of five or six people engaged in a lively discussion, surrounded by hundreds of computers. Their only tool was a simple whiteboard. To them, technology was for execution and the real value was in generating ideas and co-creating something with their fellow human colleagues.

Patrick Hunger concludes that "When we say that we are a tech company at heart, we mean that technology is the primary instrument for us to put into practice human skills. Technology amplifies our organizational capabilities to the point that size is no longer a limitation. As humans, we aspire to deliver unparalleled market access and services to clients and to become the most professional and profitable facilitator in capital markets, and we are enabled by our confidence in technology."

We hear about technological singularity, the point in time when AI surpasses human intelligence, fueling all kinds of doomsday scenarios. But what if the actual singularity was something more akin to what we witnessed at Saxo – the moment where technology only serves as a catalyst, leaving us with a renewed confidence in our own human ingenuity. And a whiteboard to express it.

The car as a point of sale and the role of automotive banks in future mobility

ZHE HU | Associate Consultant, Capco
GRIGORY STOLYAROV | Senior Consultant, Capco
LUDOLF VON MALTZAN | Consultant, Capco

ABSTRACT

The article provides a business perspective on mobility and highlights the interdependence between mobility as a service, connected vehicles, and autonomous driving. We make assumptions on what future mobility use cases might look like and how they would transform the business models of car manufacturers, dealerships, and automotive captive banks. For captive banks, we provide some ideas for staying competitive by adapting new innovative approaches and making better use of customer data, which is undoubtedly their most important asset.

1. INTRODUCTION

Several months into 2018, and the developments in the automotive industry and the broader mobility market are more exciting than ever. Autonomous vehicles are being tested on the streets of cities around the world, from Gothenburg in Sweden to Suzu in Japan. Every major car manufacturing brand is investing in a proprietary connectivity platform to power their own breed of connected cars and, in the process, slowly morphing into an increasingly agile technology company. Ride- and car-sharing demand is reaching new heights – six million regular users according to one recent study.[1] Some of the more progressive municipalities are thinking about the benefits of having an end-to-end digital mobility platform that aggregates all modes of transportation including self-driving robo-taxis, ride-sharing, and whatever else is at the core of an urban ecosystem. Electric vehicles are becoming more affordable and far-reaching, with the most recent one being the 2018 Bolt EV by General Motors. The rapid development of mobility economy is taking place along the three vectors – mobility platforms, connected vehicles, and autonomous cars. In this article, we look at each vector and conclude with a section focused on the financial arm of car manufacturers – automotive captive banks – and approaches for them to secure their competitive position amidst the fierce competition from automotive fintech companies.

2. MOBILITY AS A SERVICE (MAAS): OWN YOUR JOURNEY, NOT THE VEHICLE

Disruptive ideas and technologies, combined with economic trends and new consumer demands, are changing the face of mobility. All over Europe, innovative mobility startups and services are being founded, and many technologies as well as services are being tested.[2] This creates a new range of opportunities for the multiple players in the mobility and transport markets, such as car manufacturers and their captive banks, as mobility increasingly becomes intertwined with consumption. For example, new technologies free up time for drivers as less time behind the wheel implies more time for work, leisure, and infotainment consumption activities. These emerging mobility ideas and technologies can be summed up in one concept – "mobility as a service" (MaaS). Through MaaS, different types of transport methods and services are integrated through a mobility platform that provides mobility services for customers' needs from a single hub (i.e., website or application).[3]

MaaS providers offer their customers different mobility services based on advanced technologies and varying transportation parameters, such as shared vehicles (i.e., cars or bicycles), access to public transport, hailing taxis, or car rental services.[4] Notable examples of ride hailing services are Uber or Lyft.

The main consumer benefit from MaaS lies in the single mobility platform, which facilitates payments via a single payment system in contrast to previous multiple payments and ticketing operations. MaaS offers its users an enhanced value proposition by, for example, simplifying payments for traveling or removing unnecessary organizational aspects of journeys.[5]

The concept of MaaS has, therefore, led to new business models and methods of organizing travel and transport. MaaS provides multi-faceted business models and opportunities for transport providers, including a larger customer base, as more users have access to the service, as well as access to information (data) regarding travel behavior, which in turn can be analyzed to identify unmet demand. A central goal of MaaS is to create a viable alternative to the use of privately owned vehicles without compromising on convenience, sustainability, capacity, costs, and traffic reduction.[6]

Examples of such mobility platforms are the Swedish provider Drive Sweden[7] and the Finnish provider Whim.[8] Both offer their customers an end-to-end approach, meaning that travelers can book and pay for a flight, train, or car from one single platform. In contrast to pay-as-you-go services, such as DriveNow[9] or car2go,[10] these new MaaS concepts allow users to pay a monthly membership fee as well as use the pay-as-you-go alternative. The Whim business model is based on users paying for different monthly mobility plans in return for access to mobility services, such as bicycles or local transport. Moreover, the basic subscription package can be upgraded through additional packages at a higher price, such as a subscription for a car. A central competitive advantage of MaaS platforms is that they

[1] Bert, J., B. Collie, M. Gerrits, and G. Xu, 2016, "What's ahead for car sharing? The new mobility and its impact on vehicle sales," Boston Consulting Group
[2] The Mobility as a Service (MaaS) Alliance, https://maas-alliance.eu
[3] Definition of MaaS from MaaS website https://maas-alliance.eu/homepage/what-is-maas/
[4] ibid
[5] Ibid.
[6] ibid.
[7] What is Drive Sweden, https://www.drivesweden.net/
[8] What is Whim, https://whimapp.com/
[9] What is Drive Now, www.drive-now.com
[10] What is Car2go, www.car2go.com

can use the existing transport infrastructure of other mobility providers, such as train companies, car fleets, and public transport and create a single hub or point of access for customers.

3. MAAS TRENDS AND CHALLENGES

Recent studies have shown that MaaS is subject to several trends, which are likely to change the face of mobility:[11]

- **New technologies:** Autonomous and connected vehicles will allow for smoother and more efficient driving, which in turn leads to a reduction in fuel consumption (gas versus hybrid versus electric).
- **Emerging services:** New mobility services will be provided in both public and private transport sectors. Urban transport services will increasingly be subject to citywide mobility platforms, which enable users to have access to autonomous public transport, autonomous taxis, and car hailing services. Mobility experts have predicted that digital mobility platforms are going to integrate all transportation methods and services and will become the centerpiece for the approach to tackling the many challenges of urban mobility.
- **Increased safety:** Manufacturers and mobility providers are working to meet the rising consumer demand for more safety in traveling by developing technologies through collected data to prevent and reduce the amount of traffic accidents in both private and public transport. As an example, connected autonomous vehicles may be able to communicate with each other in order to predict dangerous traffic situations.
- **Reduced traffic congestion:** MaaS will help reduce traffic congestion in urban locations, and as a result contribute to an increase in available parking space and a decrease in tailpipe pollution (further decreased by electric or hybrid vehicles)
- **Vehicle sharing:** MaaS services rely on consumers sharing the transportation infrastructure. There will be an increasing shift towards mobility infrastructure owned by MaaS providers. This in turn implies that manufacturers have to reassess who their future customers will be (car and bicycle fleets owned by MaaS provider versus privately owned vehicles).
- **Walking:** As city planning and urban mobility concepts evolve, more consumers will cease to rely on vehicles.
- **Ownership of MaaS:** Private companies will own and provide MaaS platforms, while cities and governments are responsible for providing a legal (e.g., authorizing testing of autonomous vehicles in traffic) and economic (e.g., access to finance/ subsidies) environment for them to flourish in.
- **Cities as drivers of MaaS:** Studies have shown that MaaS is not only a theoretical concept. Currently, cities at the core of MaaS have shown that testing of autonomous vehicle models has taken off. Examples include Singapore having AV and taxi trials since 2015; Helsinki testing AV shuttle and bus services; Wuzhen field testing of autonomous cars since 2016; and Suzu testing in public road trials since 2015.

However, the above trends are also facing some resistance:

- **Unwillingness to share:** Many consumers are still unwilling to share vehicles with others. This is due to several reasons, such as: (1) privacy – unwillingness to share a ride with total strangers, (2) habit – consumers are used to owning their vehicle (vehicle ownership is still associated with a sense of freedom, flexibility, and prestige), (3) dependency – unwillingness to be dependent on a MaaS provider or on a public institution, and (4) security – fear of crime and harassment.
- **Public buy-in:** There is still very little information available to urban dwellers about advantages of MaaS.
- **Established mobility providers versus single mobility platform:** These concepts are still being tested and are often only available in certain cities (i.e., pilot projects). This prevents broad consumer adoption.

[11] Lang, N., M. Rüßmann, J. Chua, and X. Doubara, 2017, "Making autonomous vehicles a reality, lessons from Boston and beyond," Boston Consulting Group

4. IMPLICATIONS FOR MANUFACTURERS AND CAPTIVE BANKS

Based on the aforementioned trends and challenges brought about by MaaS, we believe that there is a high likelihood that direct sales of vehicles to individual owners will fall. Furthermore, as more people choose walking and shared alternative transportation methods (bicycle or autonomous public vehicles), demand for vehicles will fall even further. However, this is not necessarily bad news for manufacturers and captive banks, as MaaS will provide a range of alternative markets and business opportunities, such as more (autonomous) car fleets owned by governments or private firms. In addition, these organizations will have access to more, and different kinds of, data, which, with improved analytical tools, could create new business opportunities.

5. REAL CONNECTIVITY MEANS NEVER GOING OFFLINE

Together with MaaS, a connected vehicle – one equipped to connect to other vehicles, devices, networks, and services inside and outside of the automobile – offers a glimpse into what mobility could look like in the future. Such changes will result in a fundamental disruption in the automobile industry and challenge our perspectives on what mobility and vehicles actually are.

Having a connected vehicle can significantly improve the traveling experience of passengers, both inside and outside the vehicle. This improvement can be made in terms of vehicle maintenance, time management, safety, and infotainment control.

A connected vehicle would have navigation tools that help the driver choose the most efficient route that minimizes traveling time, fuel consumption, and toll gates payment based on real-time road and weather conditions. It would also be able to collect user field data on vehicle usage and performance and recommend suitable maintenance actions, greatly reducing the instances of unexpected vehicle breakdown and the associated frustration and inconvenience. As such, with better navigation and predictive maintenance functions, drivers would be able to reduce unnecessary time as well as costs associated with vehicle usage.

Real-time emergency calls and pre-collision warnings can also become features of the connected vehicle, significantly improving safety on the road and that of the driver and passengers. Drivers and passengers of such vehicles can also link their mobile phones and other smart devices to the vehicle to listen to news, music, access email, and take phone calls while driving, making traveling more interesting and productive than before. But hold on, this is already reality.

According to Forrester Research,[12] in the future mobility apps will be able to "provide a range of services such as mileage tracking, parking location reminders, diagnostic assistance, crash alerts, and remote control. And some, such as Verizon Hum, provide roadside service and hands-free calling. Tablet-based connected car systems such as Garmin's DriveAssist, Rand McNally's OverDryve, and Parrot's Asteroid line offer a range of features such as navigation, music, voice control, speakerphone, dash cam, backup cam, and driver warnings for collision avoidance or lane departure."

Beside the added convenience and improved safety for the driver and passengers, there exists a huge potential for the automotive industry and related stakeholders. Everything starts with the automobile manufacturer, who needs to design and build the hardware of the vehicle so it can be equipped with the necessary sensors and software to connect it to other devices. Then, the software engineers can work their magic and turn the vehicle into a "data harvesting machine."[13] The data collected when the vehicle is in action can then be mined, used, and monetized by the many stakeholders in this industry.

The "original equipment makers" (OEMs) and dealers can make use of the user field data to monitor the performance of the vehicle and provide diagnostic and preventive maintenance. Such data can be used to make warranties more effective and more tailored to the uses of individual vehicles and customer needs. For example, automatic scheduling of maintenance appointment can be made based on actual vehicle usage. This scheduling can also be linked to the OEMs, helping them with their inventory management and letting them know the type and number of component parts to prepare for a certain customer at a specific time. Information providers, such as radio stations, can provide infotainment and traffic conditions to the vehicle based on the preferences and location of the driver and passengers, collected over time via vehicle usage.

Connected cars also have the potential to disrupt the auto insurance market by providing pay-as-you-drive insurance. This type of insurance can achieve a better segmentation of customers as well as better alignment of insurance and the risks involved.

"According to Forrester Research, in the future mobility apps will be able to provide a range of services such as mileage tracking, parking location reminders, diagnostic assistance, crash alerts, and remote control."

Retailers can also make use of such data to bring targeted advertising and offerings to the driver and passengers based on their personal preferences and locations straight into the car, and also provide on-demand information about the retail stores in the vicinity.

Last, but by no means least, the data can also be used by roads and traffic authorities, police, and hospitals. Connected vehicles would enable emergency, distressed, and breakdown calls, as well as vehicle data based road maintenance. In addition, better speed monitoring and road toll systems can be set up and traffic flows can be better managed.

Along with a huge potential in this emerging connected vehicle movement come a few risks that cannot be ignored. Certain necessary changes need to be made in order for the connected vehicle ecosystems to function and flourish.

It could be said that the most important issue with the connected vehicle is data protection, since the misuse of such data could cause unimaginable damage to end-users and the society at large. The information security framework, as set by the U.S. National Institute of Standards and Technology (NIST),[14] which covers confidentiality, integrity, and availability of data, is a good starting point to tackle this issue.

[12] Gillett, F. E., 2016, "The retrofit future of the connected car," Forrester
[13] Kaminska, I., 2017, "Your car as a data harvesting machine," Financial Times, November 24, http://on.ft.com/2FbDOj7
[14] The National Institute of Standards and Technology (NIST) has core competencies in Measurement science, rigorous traceability, development and use of standards, https://www.nist.gov/

The car usage data, as well as the data on drivers and passengers, need to be properly managed. Given its huge commercial potential, it could be hacked, misused, altered without permission, or lost. When such things occur, the safety of the driver and passengers could be compromised, and the owner of the data could face financial, regulatory, and reputational damage. Hence, it is essential that automakers, technology firms, and the government work together to set laws and regulations in place to protect the data that is generated by connected vehicles. They need to agree on what data can be collected and whether consent is needed from the driver and passengers.

How this data is used is also an issue that needs to be considered carefully. For example, if the connected car detects that the driver regularly visits doctors and pharmacies, should this information be given to medical providers in case the driver has an emergency medical situation? Should this information also be given to the car insurance provider who might need to reconsider the insurance plan for the driver if his/her medication or medical condition increases the risks of road accidents? For questions like these, there are no clear answers and it is up to the stakeholders involved to come up with solutions that are balanced, fair, and beneficial.

6. HANDS-FREE BOOSTS PAID-FOR SERVICES

An extension and further evolution of connected cars is the autonomous car: in addition to being connected to other devices, infrastructures, and networks, an autonomous car requires no human driver. Currently, the autonomous car is still in development. Since 2010, technology firms and automakers have invested more than U.S.$ 111 billion[15] to fund the research and development of both semi-autonomous cars, which operate with driving assistants, and autonomous cars, which operate completely on their own.

On top of the aforementioned market disruptions caused by connected cars, the introduction of autonomous cars will bring about even bigger changes in society, especially in retail, advertising, and traffic management. This is because with assisted driving, or self-driving functions, people would have the capacity to turn their attention to other things than the driving wheel while in the car. This is aptly described by a research done by Gartner Inc.,[16] which compares the automobile evolution to "embracing the automobile as a critical element in users' digital lifestyles. Connected drivers are ultimately connected customers and consumers who increasingly have a desire for consuming, creating and sharing digital content in all situations – including when being mobile in an automobile." The combination of autonomy and connectivity will create a "third space"[17] that is neither home nor work, where people will have the time and opportunity to engage in activities of their choice. This is where retailers and advertisers have the potential to transform an autonomous vehicle into a moving retail store or a digital experience center. Such a vehicle, with digital technologies and attention from passengers, allows businesses to shape passengers' buying habits much more than what cars can currently do, and will transform the entire mobility concept and experience for passengers. With the self-driving functions, autonomous vehicles can be used as privately-owned cars, or robo-taxis and robo-buses. More people would be using vehicle sharing services and they would no longer need to spend time to look for parking space when they have arrived at their destination, as their cars can simply drive away to find a parking space and come back to fetch the passengers when needed. This greatly reduces road and inner-city congestion and results in more efficient traffic and also cleaner air.

In order for this automobile revolution to happen, an entire ecosystem needs to be set up. All stakeholders must work together for autonomous vehicles to be running at full speed. First, there must be support at the governmental level to ensure necessary regulations and infrastructures are in place. Local authorities, such as road, traffic, and city planning authorities, must collaborate to accommodate the introduction of autonomous vehicles. For example, new traffic rules might need to be drafted. Also, city roads and landscapes would need to be modified to allow for less parking space and more charging points. Governments should also invest in high-speed internet connections to enable better connectivity of the autonomous vehicles to other devices and systems. Further, as noted earlier, governments should create laws that protect privacy of the data generated by the passengers of autonomous cars.

[15] Kässer, M., T. Müller, and A. Tschiesner, 2017, "Analyzing start-up and investment trends in the mobility ecosystem," Mckinsey & Co.

[16] Ramsey, M., and J. F. Hines, 2016, "Master the four stages of connected-vehicle evolution to lead in the renaissance of the automobile," Gartner

[17] Mckinsey, 2016, "Car data: Paving the way to value-creating mobility, perspectives on a new automotive business model," Mckinsey & Company, http://bit.ly/2F3zOxO

Second, technology firms and automakers must be willing to invest in the research and testing of the autonomous vehicles features, their safety on the road, and data security in the car. This would be reflected in the number of new patents generated in relation to autonomous vehicles, the number of partnerships technology firms have with automakers, and in the amount of investments injected in developing autonomous vehicle technology.

Finally, it is important to have the acceptance from consumers – the ultimate end-users of autonomous vehicles, whose data can be mined and monetized. Consumers must be willing to, and comfortable with, sharing data about their habits and preferences with entities, such as automakers, technology firms, and possibly retailers and insurance companies. It is only with the buy-in from all three stakeholders that the ecosystem for autonomous vehicles can grow and flourish.

Currently, many countries are partnering with automakers and technology firms to test the feasibility and viability of autonomous cars. Acceptance rate of end-users differs from country to country, although consumers are in general very open to connected cars, which are already in the market, and autonomous car sharing, which is still in development. This is because the cost of transport will be significantly reduced, and the convenience and ease of traveling will be markedly increased with the introduction of autonomous vehicles. The cost of transport is reduced through the elimination of the driver, higher frequency of car sharing, and greater utilization rates of the vehicle.

The added convenience of traveling is a result of the possible "on-demand" vehicle, which eliminates the time to find a parking spot or to wait for public transport or regular taxi. Research has shown that such benefits of autonomous cars are more pronounced in cities where population is dense and aging, and the infrastructure and public transport system are reaching their capacity. Autonomous cars can help relieve the strain felt by the current transport system through providing additional options for traveling. They can also improve safety on the road and convenience for individual passengers, especially if they are elderly or have restrictions with taking public transport.

Not only will autonomous cars benefit individual travelers, they can also make the entire area where they operate better by increasing road safety (which is estimated to result in the elimination of around 3 million accidents a year in the U.S. alone)[18] and lowering pollution.

There is no doubt that autonomous cars will be the future and that future is coming sooner than we think.

7. NEW BUSINESS MODELS FOR NEW MOBILITY

We have so far discussed the established mobility trends and their socioeconomic implications for consumers and the projected impact on the automotive industry and major stakeholders in the public sector. In this section, we will discuss the automotive brands themselves and point out the strengths and weaknesses of their market position and suggest the vector of change.

All players are fully aware that finding solutions for new mobility concepts is not merely a nice-to-have, but essential for survival. The solutions are, however, not possible without rethinking the existing relationship between the manufactures, dealerships, and automotive captive banks.

To start with, it should be recognized that, at least in Europe, consumers still prefer that dealerships configure their car and request a financing offer. However, Elon Musk has shown that this does not have to be the case in the future. Configuring and buying a car online has worked well for Tesla, so why pay dealerships for their dispensable service?

The digital sales channel provides the inquisitive and demanding customer with complete information about the car, as well as the possibility to compare conditions and financing options from many providers, from the comfort of their home. Hence, the only reason to go to a dealership would be the test drive.

Captive banks are very good at providing flexible financing both to dealerships and end-customers. They have promptly responded to the customer demand for "using" instead of "owning" and are important providers of leasing schemes, along with non-captive specialists, such as Santander. However, are captives ready for the paradigm shift in the direction of the connected car and

[18] Collie, B., J. Rose, R. Choraria, and A. K. Wegscheider, 2017, "The reimagined car; shared autonomous and electric," Boston Consulting Group

the instant payment platform concept, instead of batch processing and T+2 payments status quo? The former is inevitable, when the car ultimately turns into a point of sale, to a smartphone on wheels, where you can buy car apps instead of iOS or Android apps.

We have already mentioned the pay-per-use services that will exist next to pay-as-you-go or pay-how-you-drive services. They require a seamless instant payment infrastructure and activation of the purchased service on-the-fly. The payment infrastructure should be robust to allow for a high volume of micropayments with very low transaction costs. Today, the captives in Europe are not yet capable of performing this task.

Finally, in the future, the connection between the customer and the manufacturer will not break after the purchase of the car is accomplished. The brands that will be successful in becoming digital technology companies will be able to create an ecosystem of connected cars that can be accessed remotely and can exchange data between themselves and the cloud. Data that these cars generate will become the single most important asset for the car manufacturers. Based on this data, customer experience is improved, and hence the better product can, and will be created. The security of autonomous fleets will also depend on the quality of the connected car platform. It can be observed in the market that the manufacturers are the best at coping with platform challenges, whereas captives and dealerships are still some way behind. What steps can captives and dealers take to catch up?

7.1 Buying a car at the bank

Interesting cases of proactive banking were mentioned in the automotive finance study by Nextcontinent.[19] According to the study, a number of French banks have begun selling cars directly to customers. Using their proximity to customers, these banks overtake captives and sign a contract with a customer before they enter the dealership. Similar strategies can be applied by captive banks themselves. Having proximity to both car manufacturers and customers, captives can use their online channel to offer both a car and a financing and insurance package, thus making a visit to the dealership unnecessary.

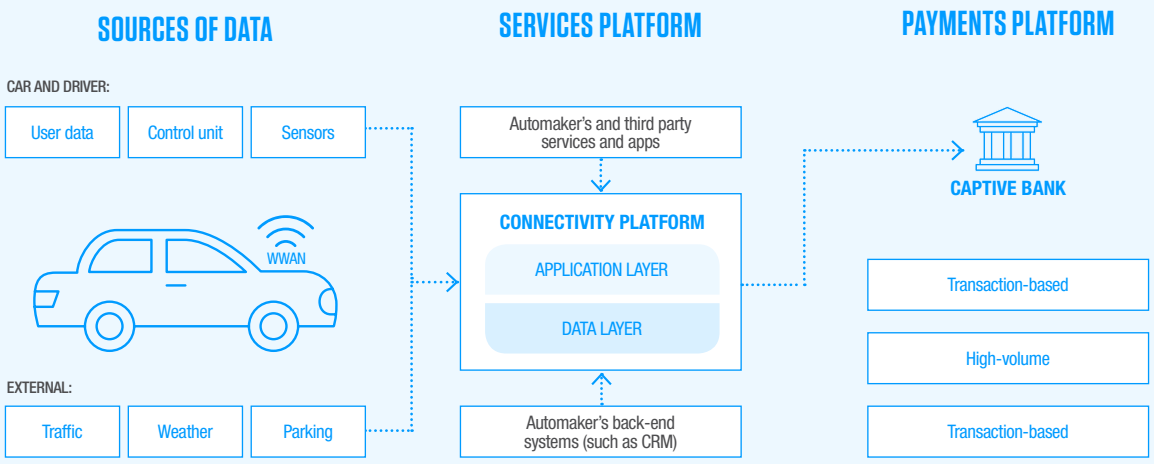
7.2 Product innovations and utilizing the dealerships network

A good example[20] of this approach is a German bank that works closely with dealers to capitalize on their customer base. The car price tag at the dealer bears a QR code. Customers can scan the code and be directed to the bank's financing landing page, which is prefilled with the particular car details. Customers can then obtain a complete online credit check by entering their personal and financial details. The client data is then fed back to the dealership, which can follow up with the client on their purchase intention, as can the bank. This example shows how a bank can integrate customers' buy-signals into their digital workflow and simplify the purchase and credit check process for the customer.

[19] Nextcontinent, 2016, "Automotive finance study 2016; the European market and its future challenges," http://bit.ly/2siGwc7

[20] Ibid.

Figure 1: When captives, manufacturers, and consumers come together



Source: Kilian, R., C. Gauger, J. Stein, and M. Scherer, 2017, "Connected vehicles and the road to revenue," BCG

Another approach[21] from Germany is transmitting customers' personal details to the bank by letting them scan their ID cards using an ID scanning device at the dealership. The data is then stored both with the bank and the dealership.

In the above examples, we can see the smart realization of the credit check process by supporting the customer in their task of obtaining a financing offer and offering them a single point of contact instead of two. In addition, there is always a small percentage of early adopters of new services and devices, who are likely to become curious about innovations by the dealer. They will be the first to test and use the new services and spread the word if they like the experience.

7.3 Knowing your customers' data

As has been shown, customer data is key to boosting revenues. It should be captive banks' top priority to collect, process, and utilize customer data by creating automatic workflows based on recognized behavioral patterns and customer lifecycle events. There are numerous ways in which customer data can be collected and legitimized for corporate use. Data can be gathered both in the pre-sale and in the car-ownership phases. While the pre-sale data mining has been well exploited in the industry, the car-in-use data collection is a new terrain for most brands, made possible by recent technological advances in automobiles.

Tesla is a well-known player in the data collection game, as it is constantly and relentlessly tracking the usage patterns of drivers of its growing fleet and offering them regular OS updates in exchange for this, making the driving experience better as a result. Some of the more advanced car rental firms are also starting to use the location and fuel status data. By using WWAN (wireless wide area network), Bluetooth, and other protocols, these companies are able to access car data remotely and analyze them in real time. However, client personal data underlies a stricter regulation and cannot currently be used for modeling purposes.

It would be another natural step for captives to work more closely with their car manufactures to utilize the full scale of car usage data regarding specific customers. This data would include car status, personal, and financial customer data. Customers will allow access to the full set of data, if they receive a new level of customer service, infotainment, and security in exchange. Collection of this data is an inevitable milestone on the way to realization of a customer-manufacturer-bank platform. Figure 1 illustrates what that end-game scenario could look like.

[21] Ibid.

7.4 Investing in digital services

In the past, captives have done well in supporting car manufacturers. They enhanced the value chain and offered innovative financing solutions. But today, captives need to move away from reactive, product-centric operating models, and consider offering a customer-centric portfolio of digital mobility solutions tailored to the fast-changing needs of car buyers.[22]

Today, over 46% of vehicles in Germany are financed through captives, which leaves them with an incredible amount of customer data. Car manufacturers and dealers have little access to this data. However, it must be said that captives themselves make little use of the precious asset they possess. Innovation rate has been traditionally low with captives and soon the growing tech-savvy competition will threaten their positions, especially in the wake of PSD2.

However, if captive banks are up for the fight, the right direction would be the creation of a holistic portfolio of digital products, centered around customer needs. Besides the classic financing and leasing options, mobility solutions related to, for example, car sharing, car hailing services, or autonomous driving will help them stay competitive. A car with enabled connectivity can and should receive a unique digital identifier, with which the active elements of infrastructure can instantly identify it. By virtue of this ID, the customer will no longer need to leave a car to pay for fuel or parking. The car number plate coupled with the wireless ID information is enough to lead to the digital wallet connected to the customer's bank account. Captive banks need to recognize this tendency early and make mobility related infrastructure solutions part of their digital services portfolio. Automotive companies have to transform into mobility solution providers – and their captives are best equipped to deliver these products to their customers.[23]

8. OUTLOOK

In the middle- to long-run, connected cars will be as widely used as smartphones. We will see the establishment of several competing ecosystems that will further drive the evolution of car-based applications and services. It will not be surprising to see more successful connectivity platforms being based on an open source code – a strategy proven in many other industries. Autonomous cars will soon appear on the streets of most major cities. It is not clear yet how consumers will use them – be they privately owned or shared vehicles, private robo-taxis, or municipal robo-buses. What is clear, however, is that connected autonomous fleets will grow and benefit many stakeholders – from delivery services to logistics companies, from small communities to megapolises. The new lucrative market for car applications, with and without interaction with other infrastructure elements, will soon be filled by all kinds of fintechs, developers, and enthusiasts that will deliver services that we cannot even imagine today. By that time, instant payment platforms will have long become a reality, but it cannot yet be said who will take the lead – fintechs, captives, or non-captive banks.

[22] Coccorullo, S., 2016, "Unlocking the hidden potential of automotive captive banks," Oliver Wyman
[23] ibid.

Fintech and the banking bandwagon

SINZIANA BUNEA | University of Pennsylvania
BENJAMIN KOGAN | Development Manager, FinTxt Ltd
ARNDT-GERRIT KUND | Lecturer for Financial Institutions, University of Cologne
DAVID STOLIN | Professor of Finance, Toulouse Business School, University of Toulouse[1]

ABSTRACT

We examine annual report filings of U.S. listed banks to assess their attitude toward the financial technology (fintech) sector. Banks did not mention the impact of fintech on their business until the 2016 filing season, when 14 banks, or 3% of the total number of filers, did so. In 2017, this number skyrocketed to 66 banks, or 14% of filers. These filings prove to be a rich source of data about banks' perceptions of fintech. Further, compared to survey data, the information has the advantage of being management-certified and not anonymous. We analyze what banks say about fintech in their annual filings and find that they are largely concerned about fintech's impact on deposits, lending, and payments business, and about the proposed bank charter for fintechs. Banks are much less worried about cryptocurrencies, blockchain, and competition from "Big Tech." There is also evidence of banks being influenced by what their peers are saying, and even copying peers' disclosures verbatim.

[1] We are grateful to Yuliya Snihur and Maxim Zagonov for suggestions, and to Jasper Ginn for help with data. All errors are ours.

1. INTRODUCTION

In the post-dotcom-boom era, it is rare that a business sector, let alone one as large and long-established as banking, exhibits a fivefold year-on-year increase in anything. Yet, this is just what happened over the 2016-2017 period with U.S. banks, whose official notifications to their investors about the impact of fintech on their business grew 371%. What does this remarkable fact mean, and what do we learn from it?

In this article, we focus on the text of banks' disclosures about fintech in their Form 10-K annual reports. In one year, the amount of textual data increased dramatically, yielding a number of intriguing insights into an important industry's official reaction to digital disruption.

First, banks that mention fintech overwhelmingly view it as a threat rather than as an opportunity (though many more do not mention fintech at all, at least for now). Further, even among those that refer to fintech, many do not state the exact nature of the threat. For those banks that do, deposits, lending, and payments are deemed to be the most vulnerable lines of business. Only six banks discuss their strategy with respect to fintech competition, with four stressing potential partnerships with fintech firms. We also report that banks that refer to fintech are somewhat more likely than those that don't to be already engaged with the fintech sector in a meaningful manner, and to be involved in industry consolidation.

The sudden increase in the use of the term fintech in banks' 10-K forms makes one wonder what has prompted it. One reason could be the fact that the Office of the Comptroller of the Currency (OCC) has raised the possibility of a special bank charter for fintech companies, which appears to have alarmed 11 banks. We also present evidence that some banks simply copy their peers in how they refer to fintech, and point out a geographical pattern in the way references to fintech have spread across banks. Finally, we find that when compared to their concerns about competition from fintech companies, banks seem to be less worried about the competition from Big Tech.

2. OVERVIEW OF THE DATA

Table 1: Determinants of banks' fintech mentions

PANEL A: DESCRIPTIVE STATISTICS

	MEAN	MEDIAN	SD	MIN.	MAX.
log(assets)	7.979	7.632	1.593	4.785	14.728
ROA	0.008	0.009	0.004	-0.030	0.030
ROE	0.074	0.080	0.079	-1.251	0.310
CEO age	59.670	60.000	6.690	36.000	86.000
CEO compensation	1,877,281	807,713	2,902,329	26,804	27,236,892
Long-term compensation	0.284	0.226	0.278	0.000	1.000

There is a substantial finance literature on the informativeness of textual disclosures by companies in general, and in the risk factor disclosures of their 10-K filings in particular [Campbell et al. (2014)]. To conduct our analyses, we look for the text "fintech" or "financial technology" in 10-K forms filed in the 12 months ending 31 December, 2017 with the Securities and Exchange Commission (SEC) by publicly traded U.S. bank holding companies (which we define as corporations whose Standard Industrial Codes, or SICs, range from 6021 to 6036 according to the SEC's EDGAR portal). We retain only disclosures addressing the impact of fintech on the bank's business,[2] and henceforth refer to these cases as "bank fintech mentions". Our 2017 sample consists of 66 banks with such mentions (these are listed in Appendix A), up from 14 in 2016 and zero in the preceding years.

"The number of bank annual reports mentioning fintech competition grew from 0 in 2015 to 14 in 2016 and to 66 in 2017."

[2] This excludes one bank that mentions fintech in an executive's biography and another that mentions a loan portfolio acquired through a fintech company.

Table 1: Determinants of banks' fintech mentions

PANEL B: RESULTS OF PROBIT REGRESSIONS

	LOG(ASSETS)		ROA		ROE		CEO AGE		CEO COMPENSATION		LONG-TERM COMPENSATION	
	coef.	p-value	coef.	p-value	coef.	p-value	coef.	p-value	coef.	p-value	coef.	p-value
MODEL 1	0.2651	0.0000			-0.6100	0.5195	0.0075	0.5279				
MODEL 2	0.3567	0.0000			-0.7188	0.4386	0.0062	0.6056	0.0000	0.3561	-0.1575	0.7062
MODEL 3	0.2665	0.0000	-10.0393	0.5629			0.0074	0.5344				
MODEL 4	0.3642	0.0001	-14.7381	0.4072			0.0060	0.6156	0.0000	0.3407	-0.1731	0.6799

To provide a better understanding for what bank characteristics are linked with the likelihood of a fintech mention, we conducted probit regressions, as reported in Table 1 (Panel A summarizes the explanatory variables, and Panel B shows the estimation results). The dependent variable is 1 for each of the banks that mention fintech in 2017, and 0 for other filing banks. Our explanatory variables are bank size (the logarithm of bank assets) and profitability (ROA and ROE), as well as CEO characteristics (age, total compensation, and the proportion of long-term compensation) in 2016. The table shows that only bank size is a significant predictor of whether a bank will mention the fintech threat: larger banks are more likely to do so.

With respect to the content of the banks' filings, of the 66 fintech-mentioning banks, 20 use only the term "fintech,"[3] 30 use only the term "financial technology," and 16 banks use both terms. All but two of the banks include these mentions in Item 1 (Business) or Item 1A (Risk factors) of the filing.[4]

Among the 66 filings, 58 mention fintech as a competitive threat, six mention it as both a threat and an opportunity, and two only as an opportunity.[5]

Of the 64 banks that see fintech as a threat (including the six that also see it as an opportunity), 23 do not provide further information, simply mentioning fintech as part of a list of competitor types,[6] or separating it from other competitors in a statement such as "in addition, financial technology, or fintech, startups are emerging in key areas of banking."

Such statements are often boilerplate. For example, seven banks opine that fintech "made it possible for nonbanks to offer products and services traditionally provided by banks" and three state that "technology and other changes, including the emergence of "fintech companies" are allowing parties to complete financial transactions through alternative methods that historically have involved banks." However, there is also variation and richness in the fintech mentions that allows us to examine banks' attitudes toward fintech, and we turn to this issue next.

[3] This includes "fintech" (3 banks), "fintech" (4 banks), and even "fin-tech" (1 bank).
[4] The two exceptions include it in Exhibit 13, containing additional sections of the bank's annual report to investors, such as the letter to shareholders.
[5] It is notable that both of the banks that mention fintech as an opportunity recently became involved in the fintech space – Citizens Financial through a partnership with Bottomline Technologies, and Live Oak through its fintech venture capital arm.
[6] For the record, of the 23 banks, 14 mention fintech in a list (ranging from 3 to 18 competitor types, with a mean and median of 7); in 8 of the cases, fintech is the last item on the list.

3. HOW BANKS DEFINE FINTECH AND WHAT WORRIES THEM ABOUT IT

As Bunea et al. (2016, footnote 3) discuss, the earliest known definition of fintech, attributed to Bettinger (1972), who then worked at Manufacturers Hanover Bank, remains pertinent despite the very different context of the time: "an acronym which stands for financial technology, combining bank expertise with modern management science techniques and the computer." However, while Bettinger was preoccupied with better analytics, thirty-five years on, banks' definitions of fintech are particularly concerned with product delivery, describing fintech in their 10-K filings as: "a broad category referring to technological innovation in the design and delivery of financial services and products" (Mainsource Financial Group); "[the use of] new technology and innovation with available resources in order to compete in the marketplace of traditional financial institutions and intermediaries in the delivery of financial services" (Orrstown Financial Services); or "companies that provide innovative web-based solutions to traditional retail banking services and products" (Valley National Bancorp).

While the wording of the definitions can vary, it is quite clear that by fintech, these days banks mean potential digital disruption of the banking sector, as well as the companies contributing to this disruption. Further, the 66 banks in our sample believe that the potential impact of fintech is sufficiently important for them to start informing their investors about it. So, what is it that worries banks about fintech?

Primarily, banks are concerned about competition from fintech for their core businesses: deposits (13 banks) and especially lending (20 banks). The prominence of the latter concern is unsurprising, given the proliferation of marketplace lenders. Payments, like deposits, concern 13 banks. Money transfers are in fourth place (8 banks). Other concerns are much less common: only three banks are worried about fintech competition for money management, investment advising, or leasing, and two about mortgage lending. Account management, product access, and working capital lending are singled out by one bank each. Interestingly, only one bank mentions bitcoin, while three banks mention blockchain technology. One bank states that fintech poses a competitive threat in all areas of the bank's operations.

In addition to the above threats to the banks' product offerings, several banks discuss broader threats to the banks' operations. Thus, four banks state that trying to keep up with fintech exposes them to greater cyber-security risk, five say that it can make it harder to attract and retain customers, and one bank warns about fintech making it harder to attract and retain employees.

4. WHAT BANKS PLAN TO DO ABOUT THE FINTECH THREAT

It is reasonable to assume that prior to signing off on an annual report where fintech is newly added as a competitive factor, the bank's leadership will have discussed its strategy for dealing with it. However, most banks do not choose to share this strategy with their investors. The exceptions are below:

- **Associated Banc:** "Strategic planning remains important as we adopt innovative products, services, and processes in response to the evolving demands for financial services and the entrance of new competitors, such as out-of-market banks and financial technology firms."[7]
- **BNY Mellon:** "We are collaborating with clients and leading financial technology startups, or fintechs, to develop and integrate new solutions and services, and attracting top information technology talent through our Innovation Centers worldwide."
- **Citizens Financial:** "We are also focused on fintech partnerships that help deliver differentiated digital experiences for our customers."
- **Hamilton Bancorp:** "Hamilton is evaluating fintech companies with the possibility of developing relationships for efficiency in processing and/or as a source of loans and other business."
- **Huntington Bancshares:** "We are monitoring activity in marketplace lending along with businesses engaged in money transfer, investment advice, and money management tools. Our strategy involves assessing the marketplace, determining our near term plan, while developing a longer term approach to effectively service our existing customers and attract new customers. This includes evaluating which products we develop in-house, as well as evaluating partnership options, where applicable."

[7] While one could expect such a strategy formulation to be bank-specific, United Bancshares' text, appearing 22 days after Associated Banc's, is identical to it.

- **Pacific Mercantile Bancorp:** "Thus we have reduced and expect to continue to reduce the size of our branches and are redeploying the cost savings to expand our business development team and more actively promote our online banking."

In 2016, only two banks (Hamilton and Huntington) discussed their fintech-facing strategies, hence the year-on-year increase should be welcomed by bank shareholders. However, the disclosure by Huntington, the first-ever bank to mention fintech, remains the most extensive. One could make the argument that after (justifiably) informing investors about the potential implications of fintech on their businesses, banks should reassure them about having a plan to deal with these implications. Such text could have the additional benefit of signaling to fintech companies which banks are particularly receptive to a partnership.

5. ARE BANKS' FINTECH REFERENCES JUST WORDS?

In Bunea et al. (2016), we addressed the possibility that the threats and opportunities of digital disruption in banking may be discussed without using the term "fintech." To examine whether this is so, that study formed a control sample of banks that are of similar size to the 14 fintech-mentioning pioneers, but did not employ the term in their 2016 filings.[8] The 2016 study documented that disclosures by fintech-mentioning firms do a better job of informing banks' investors about the nature and possible consequences of digital disruption. Further, the 2016 study found fintech-mentioning banks to be more active in the fintech field than control banks.

Another year of data lends further support to the notion that banks' mentions of fintech are not (just) words. First, given corporations' reluctance to change text in their annual filings and presumed internal processes that need to be involved in such changes [Cohen et al. (2017)], the very fact that 52 banks did so in 2017, is quite striking.

Second, we once again examine fintech-related actions of our 2016 sample and control banks. As of end-2017, eight of the banks that mentioned fintech in 2016 have significant involvement in the field (through fintech acquisitions, strategic partnerships, or accelerators), as compared to their four control banks.

Third, we examine acquisitions within our sample of bank holding companies. Of the 19 banks that were acquired by other companies in the sample, four banks (or 21%) are fintech mentioners. Of the acquiring banks, six banks (or 32%) are fintech mentioners. While the numbers above are not statistically significant, they do raise the possibility that fintech-aware banks are likely to play a leading role in the consolidation of the banking industry.

6. THE OCC FINTECH CHARTER PROPOSAL

The decision by the OCC to consider giving fintech firms a bank charter appears to have prompted multiple banks to address the fintech threat in their filings. A total of 11 banks discuss the OCC announcement (only one of these banks mentioned fintech in 2016). One of the first official reactions, on 2/26/17, was Iberiabank's (it was subsequently reproduced verbatim by First Financial Corp on 3/9/17 and by Mainsource Financial Group on 3/10/17): "The federal charter would largely allow fintech companies to operate nationwide under a single set of national standards, without needing to seek state-by-state licenses or joining with brick-and-mortar banks, and may therefore allow fintech companies to more easily compete with us for financial products and services in the communities we serve."

The above was echoed by both Central Pacific and CVB Financial on 3/1/17: "Recent developments include: […] the pronouncement by the Office of the Comptroller of the Currency of a limited-purpose "fintech" national bank charter which would enable a fintech company to originate loans and access the payment system directly, without relying on third-party banks. Such a development could further increase competition in the financial services sector, including with the Company and the bank."

However, the pace of OCC announcement mentions did not pick up, and in fact the last ten fintech-mentioning banks (those filing after 3/10/17) did not mention the fintech charter at all. Subsequent years' filings will likely shed more light on just how impactful the banks deem the proposed charter to be.

[8] Eight of these banks mention competition from "non-banks" (or "nonbanks"). Note, however, that this term is quite ambiguous, as the following elaboration by Bryn Mawr Bank makes clear: "The Corporation's competitors include other community banks, larger banking institutions, trust companies and **a wide range of other financial institutions such as credit unions, registered investment advisors, financial planning firms, leasing companies, government-sponsored enterprises, on-line banking enterprises, mutual fund companies, insurance companies and other non-bank businesses**" (emphasis ours).

7. THE GEOGRAPHICAL DISTRIBUTION OF FINTECH MENTIONS

The 66 banks in our sample represent 27 different states. There is some prima facie evidence of geographic clustering: although fewer than one in seven banks nationwide mentioned fintech, both banks in Utah have done so, as well as two of Georgia's three banks, half of Virginia's six banks, and seven of Indiana's 15.

Figure 1 shows the spread of fintech mentions in the Continental U.S. Banks that first mentioned fintech in 2016 are marked with gray circles, banks that first mentioned fintech in 2017 are identified with black circles, and the remaining banks with black dots. Eyeballing the data does make it appear as though there are swathes of the country where bank holding companies do not mention fintech, as well as areas where bank mentions are concentrated.

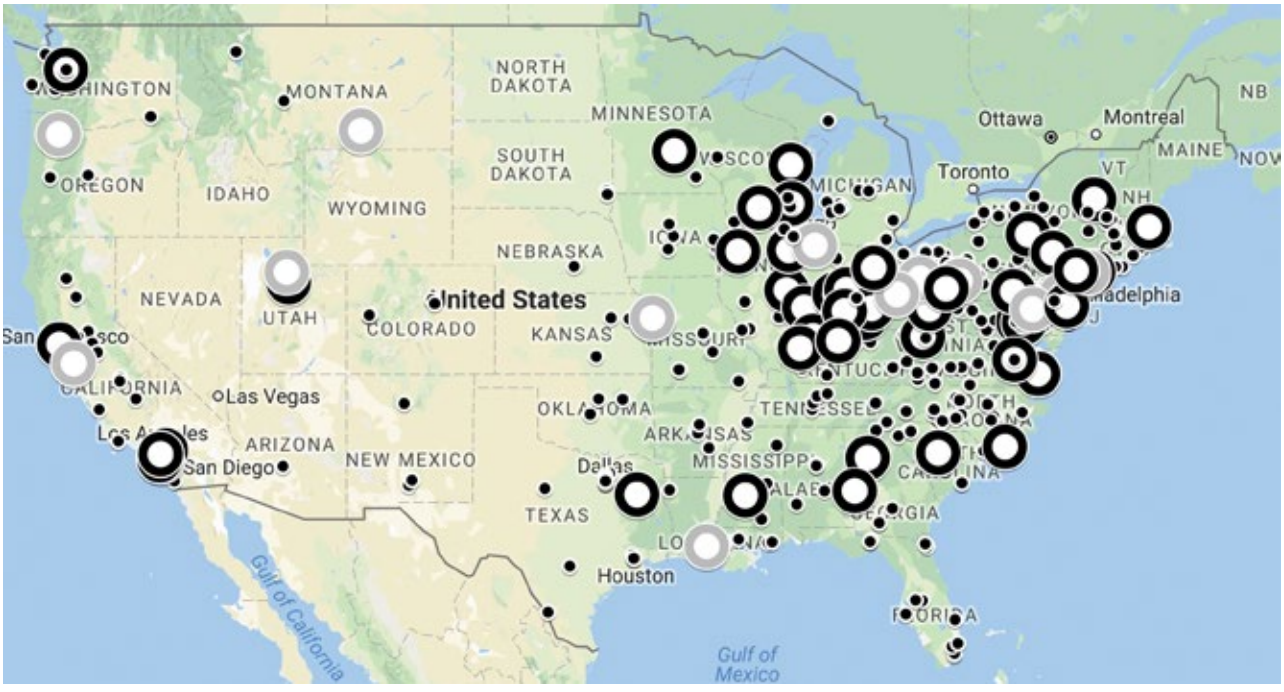
To investigate the geographic clustering hypothesis more formally, we can proceed as follows. The 14 banks that mentioned fintech in 2016 represent 11 states. Of these states' 166 other banks, 25 or 15.1% mentioned fintech in 2017. By contrast, of the 285 banks located elsewhere in the U.S., only 27, or 9.5%, mentioned fintech in 2017. According to a one-tailed Z-test for difference in proportions, this difference is statistically significant at the 95% confidence level (p-value = 0.038). In fact, given the sequence of events (a fintech mention in 2016 followed by more mentions in the same states the following year), this suggests that the increase in fintech-mentioning took place in part through geographical proximity.

8. WHY SUCH AN INCREASE IN FINTECH MENTIONS?

This is a worthwhile question, given that banks' fintech mentions came out of nowhere in 2016, and exhibited a staggering 371% increase in the following year. What explains so many banks jumping on the fintech mentioning bandwagon?

First, it is undeniable that fintech is increasingly part of a reality banks face in their strategic thinking. Given the prominence of fintech over at least the last few years,

Figure 1: The spread of fintech mentions in the Continental U.S.



Banks that first mentioned fintech in 2016 are marked with gray circles, those that first mentioned it in 2017 are marked with black circles, and non-mentioning banks are marked with black dots. An animated version of this figure showing the evolution of fintech mentions over time is available at **www.fintxt.com/s/fintech2.gif**

however, this reality alone is unlikely to come anywhere close to explaining a nearly-fivefold increase in banks' fintech mentions over the course of just one year. Even aggressively assuming that the importance of fintech doubled in a single year, we would expect to see only 14 new fintech mentioners, not 52.

Second, the OCC's announcement about granting fintech companies a bank charter was mentioned by 11 banks, so it is reasonable to assume that their fintech mentions were prompted by this announcement.

Third, like any corporation, banks can be expected to read their peers' filings. Indeed, the verbatim reproduction of newly added fintech-related text across filers that we have documented shows this to be the case. While it is implausible that every bank's decision-makers would read the filings of all other banks in the nation, banks may be more likely to read the filings of other banks in their home state. One reason for this is that same-state banks are more likely to be rivals and hence to pay close attention to each other; another reason is that same-state banks face the same regulation at the state level. Indeed, our earlier analysis shows that a bank is 15.1%/9.5%=1.59 times more likely to mention fintech in its filing if another bank in its state did so in the previous year. Taking this calculation at face value, this factor accounts for (15.1%-9.5%)*166=9 new fintech mentions. However, multiple banks could have (and indeed, have, according to cross-state text copying) been inspired by filings of out-of-state peers, which would increase the number further.[9]

Fourth, while we do not wish to flatter ourselves, it is conceivable that some banks' decision-makers may have been prompted to mention fintech either by coming across our 2016 article,[10] or by hearing about it on the Financial Times' influential Banking Weekly podcast.[11] It is unclear what figure, if any, to assign to this possibility.

Lastly, there must be other factors responsible for the rapid spread of fintech mentions, which future data and research may reveal.

9. WHAT ABOUT COMPETITION FROM BIG TECH?

While the potential of financial technology firms to disrupt banking is widely discussed in the academic and business circles, there have also been warnings about the effects of competition from information technology giants such as Amazon, Apple, Facebook, and Google [Arnold (2018), KPMG (2017), McKinsey (2017), World Economic Forum (2017)]. But, do the banks themselves voice such concerns in their annual filings?

While we did not find any mentions of "Big Tech" as such, we have identified several mentions of competition from "technology companies" (not preceded by the adjective "financial"); we have put this term in bold in the quotes below. In fact, Bank of America first mentioned competition from technology firms in March 2006 – a full decade before a bank annual report mentioned competition from fintech: "In addition, technological advances and the growth of e-commerce have made it possible for non-depository institutions to offer products and services that traditionally were banking products, and for financial institutions to compete with **technology companies** in providing electronic and Internet-based financial solutions."

The above text has stayed unchanged until the present day (with the small but noteworthy exception that "Internet" ceased to be capitalized in 2010). Years later, U.S. Bancorp introduced very similar text: "In addition, technology has lowered barriers to entry and made it possible for non-banks to offer products and services that traditionally were banking products, and for financial institutions to compete with **technology companies** in providing electronic and internet-based financial solutions."

On the 25th of February 2016, U.S. Bancorp explicitly listed technology companies among its competitors: "The Company competes with other commercial banks, savings and loan associations, mutual savings banks, finance companies, mortgage banking companies, credit unions, investment companies, credit card

[9] A more cynical take on contagion in the banking sector is expressed by Elon Musk, who among other things co-founded perhaps the most influential fintech startup of all, PayPal: "All the bankers did was copy what everyone else did. If everyone else ran off a bloody cliff, they'd run off a cliff with them. If there was a giant pile of gold sitting in the middle of the room and nobody was picking it up, they wouldn't pick it up, either" [Vance (2015)]. This point of view would explain not only why fintech mentions spread so fast, but also why there weren't any for so long.

[10] For the record, of the 14 non-fintech-mentioning banks in the Bunea et al. (2016) control sample, four mentioned fintech in 2017.

[11] Financial Times Podcast, 2017, "European banks on Brexit, Lloyds cyberattack and US banks on fintech," January, 24, http://on.ft.com/2HjqOpY

companies and a variety of other financial services, advisory and **technology companies**."

On the same day, BB&T filed the following text: "BB&T also experiences competition from nonbank companies inside and outside of its market area and, in some cases, from companies other than those traditionally considered financial sector participants. In particular, **technology companies** have begun to focus on the financial sector and offer software and products primarily over the Internet, with an increasing focus on mobile device delivery. […] Technology companies are generally positioned and structured to quickly adapt to technological advances and directly focus resources on implementing those advances."

In addition, two days later, the following text first appeared in UMB Financial's filing, and was replicated in Lakeland Bancorp's on the 15th of March: "Competition with financial services technology companies, or **technology companies** partnering with financial services companies, may be particularly intense, due to, among other things, differing regulatory environments."

Lastly, JP Morgan and CVB Financial, which elsewhere in their 10-K mention fintech competition, also address technology companies more generally, stated, respectively: "Competitors of the Firm include other banks and financial institutions, trading, advisory and investment management firms, finance companies and **technology companies** and other firms that are engaged in providing similar products and services. […] New technologies have required and could require the Firm to spend more to modify or adapt its products to attract and retain customers or to match products and services offered by its competitors, including **technology companies**" (JP Morgan) and "competition and innovation with respect to financial products and services by banks, financial institutions and non-traditional providers including retail businesses and **technology companies**" (CVB Financial).

We have also looked for explicit mentions of competition from U.S.-based Big Tech companies: Amazon, Apple, Facebook, Google (as well as its parent company, Alphabet), and Microsoft. We have only found mentions of Apple, and only in connection with its Apple Pay service[12]: "For example, consumers can now […] use electronic payment methods such as **Apple** Pay" (Citizens Holding Company). "Merchants may also continue to pursue alternative payment platforms, such as **Apple** Pay, to lower their processing costs. Any such new payment system may reduce our interchange income" (Banner Corporation, Charter Financial, and Colony Bankcorp).

"Big Tech is little mentioned in banks' filings, even though some analysts argue that it poses a greater threat than fintech."

What should we make of the above mentions of competition from technology firms? The alarm raised by analysts over Big Tech's potential to disrupt banking goes well beyond Apple Pay and it is doubtful that any traditional bank is immune from this disruption. McKinsey (2017), for example, stated that "Amazon, Facebook and Apple have all made platform-expanding moves into banking. Banks should consider the possibilities and prepare". According to KPMG (2017), "Recent quarters have seen […] an increasing number of large fintechs like Square and Klarna applying for banking charters. Yet while these growing players have made headlines, technology and ecommerce giants like Amazon, Google, Facebook and Apple may pose greater threats to the traditional banking model."

At the moment, however, concerns over such developments in banks' annual reports are muted, and far less widespread than banks' concerns over fintech. Since banks' mentions of competition from fintech started appearing in their annual reports years after bankers began discussing the topic in public, mentions of competition from technology giants may end up following a similar trajectory. However, given the speed and the scope of digital disruption in the industry, it is hard to argue that such a delay is in the interest of banks' stakeholders.

[12] Several other banks' reports mention Apple Pay as a service they provide to their clients, rather than as a competitive threat.

CONCLUSION

In the conclusion of our 2016 investigation, we listed six questions for the future about banks mentioning the fintech threat:

1. Is it that they are especially vulnerable in the face of this threat after all, and will this be reflected in subsequent poor performance?
2. Are they unusually prescient, and as such will exhibit greater adaptability and resilience, accompanied by strong financial results?
3. Will the performance of the fintech sector justify the concerns of our cohort of officially apprehensive banks?
4. Will disclosures about fintech competition continue to spread through banks' annual reports?
5. If so, to which banks?
6. Will most banks copy or adapt others' formulations, or will disclosures become increasingly informative?

While the jury is still out on the first three questions, we are now in a position to give qualified answers to the last three. Indeed, disclosures about fintech are spreading fast, and may continue to do so. In part, this spread is toward banks that are themselves involved in fintech, or are geographically close to banks that have mentioned fintech in their disclosures previously. As the spread of fintech disclosures continues, there is much copy-pasting from peers, but also an encouraging trend toward more informative disclosure.

This year's harvest of fintech-related disclosures promises to yield even richer insights. The passage of time will also shed light on a fundamental question underlying our research: does a bank's attitude toward fintech predict its success? We look forward to continuing with this fascinating line of inquiry.

References

Arnold, M., 2018, "Finance chiefs warn on Big Tech's shift to banking: Leading lenders fear internet giants will cherry-pick best parts of their business," The Financial Times, February 4

Bunea, S., B. Kogan, and D. Stolin, 2016, "Banks versus fintech: at last, it's official," Journal of Financial Transformation 44, 122-131

Campbell, J. L., H. Chen, D. S. Dhaliwal, H-m. Lu, and L. B. Steele, 2014, "The information content of mandatory risk factor disclosures in corporate filings," Review of Accounting Studies 19, 396-455

Cohen, L., D. Lou, and Q. H. Nguyen, 2017, "Lazy prices," Working Paper, Harvard Business School

KPMG, 2017, "The Pulse of fintech Q3 2017," http://bit.ly/2Aja11E

McKinsey, 2017, "The phoenix rises: remaking the bank for an ecosystem world," McKinsey & Co.

Vance, A., 2015, Elon Musk: Tesla, SpaceX, and the quest for a fantastic future, Ecco/HarperCollins

World Economic Forum, 2017, "Beyond fintech: a pragmatic assessment of disruption in financial services," http://bit.ly/2vcJKju

Appendix A: Fintech-mentioning banks

This list presents the 66 banks that mentioned fintech in 2017, by ticker symbol (in bold): **3CSBB** CSB Bancorp Inc, **AF** Astoria Financial Corp, **AROW** Arrow Financial Corp, **ASB** Associated Banc Corp, **BK** Bank of New York Mellon Corp, **BKMU** Bank Mutual Corp, **BMRC** Bank of Marin Bancorp, **BNCL** Beneficial Bancorp Inc, **BUSE** First Busey Corp, **CBSH** Commerce Bancshares Inc, **CFG** Citizens Financial Group Inc, **CHCO** City Holding Co, **CHMG** Chemung Financial Corp, **COF** Capital One Financial Corp, **CPF** Central Pacific Financial Corp, **CVBF** CVB Financial Corp, **ETFC** E Trade Financial Corp, **FBIZ** First Business Financial Services Inc, **FCFP** First Community Financial Partners Inc, **FFBC** First Financial Bancorp, **FFWM** First Foundation Inc, **FIBK** First Interstate Bancsystem Inc, **FITB** Fifth Third Bancorp, **FNB** FNB Corp, **FNCB** FNCB Bancorp Inc, **FRME** First Merchants Corp, **GABC** German American Bancorp Inc, **HBAN** Huntington Bancshares Inc, **HBK** Hamilton Bancorp Inc, **HBNC** Horizon Bancorp, **HMST** HomeStreet Inc, **IBKC** Iberiabank Corp, **INBK** First Internet Bancorp, **JPM** JP Morgan Chase & Co, **LCNB** LCNB Corp, **LOB** Live Oak Bancshares Inc, **MRLN** Marlin Business Services Corp, **MSFG** Mainsource Financial Group, **NFBK** Northfield Bancorp Inc, **ONB** Old National Bancorp, **OPOF** Old Point Financial Corp, **ORRF** Orrstown Financial Services Inc, **PACW** Pacwest Bancorp, **PEBO** Peoples Bancorp Inc, **PMBC** Pacific Mercantile Bancorp, **PNC** PNC Financial Services Group Inc, **PUB** Peoples Utah Bancorp, **QCRH** QCR Holdings Inc, **RBCAA** Republic Bancorp Inc, **SBSI** Southside Bancshares Inc, **SIVB** SVB Financial Group, **SNBC** Sun Bancorp Inc, **SNV** Synovus Financial Corp, **SSB** South State Corp, **STI** Suntrust Banks Inc, **STT** State Street Corp, **TCF** TCF Financial Corp, **THFF** First Financial Corp, **TRMK** Trustmark Corp, **UBOH** United Bancshares Inc, **UBSH** Union Bankshares Corp, **UMBF** UMB Financial Corp, **UMPQ** Umpqua Holdings Corp, **VLY** Valley National Bancorp, **WSBC** Wesbanco Inc, **ZION** Zions Bancorporation

Appendix B: Fintech mentions by state

This list shows the distribution of fintech mentioning banks by state. Each state's name is followed by the number of fintech mentioning banks, the total number of banks from that state, and tickers of fintech-mentioning banks (in blue for banks that first mentioned fintech in 2016): **Alabama** 0/6; **Alaska** 0/1; **Arizona** 0/1; **Arkansas** 0/4; **California** **6**/32: BMRC, CVBF, FFWM, PACW, PMBC, SIVB; **Colorado** 0/3; **Connecticut** 0/10; **Delaware** 0/2; **Florida** 0/13; **Georgia** **2**/3: SNV, STI; **Hawaii** **1**/4: CPF; **Idaho** 0/0; **Illinois** **3**/18: BUSE, FCFP, QCRH; **Indiana** **7**/15: FRME, GABC, HBNC, INBK, MSFG, ONB, THFF; **Iowa** 0/6; **Kansas** 0/3; **Kentucky** **1**/11: RBCAA; **Louisiana** **1**/7: IBKC; **Maine** 0/4; **Maryland** **1**/12: HBK; **Massachusetts** **1**/19: STT; **Michigan** 0/12; **Minnesota** **1**/3: TCF; **Mississippi** **1**/7: TRMK; **Missouri** **2**/8: CBSH, UMBF; **Montana** **1**/3: FIBK; **Nebraska** 0/1; **Nevada** 0/0; **New Hampshire** 0/0; **New Jersey** **4**/24: MRLN, NFBK, SNBC, VLY; **New Mexico** 0/2; **New York** **6**/29: AF, AROW, BK, CHMG, ETFC, JPM; **North Carolina** **1**/18: LOB; **North Dakota** 0/0; **Ohio** **7**/27: 3CSBB, FFBC, FITB, HBAN, LCNB, PEBO, UBOH; **Oklahoma** 0/3; **Oregon** **1**/3: UMPQ; **Pennsylvania** **5**/42: BNCL, FNB, FNCB, ORRF, PNC; **Rhode Island** **1**/3: CFG; **South Carolina** **1**/9: SSB; **South Dakota** 0/2; **Tennessee** 0/9; **Texas** **1**/16: SBSI; **Utah** **2**/2: PUB, ZION; **Vermont** 0/3; **Virginia** **3**/6: COF, OPOF, UBSH; **Washington** **1**/13: HMST; **West Virginia** **2**/6: CHCO, WSBC; **Wisconsin** **3**/10: ASB, BKMU, FBIZ; **Wyoming** 0/0.

Can blockchain make trade finance more inclusive?

ALISA DICAPRIO | Head of Research, R3
BENJAMIN JESSEL | Fintech Advisor to Capco

ABSTRACT

There is little doubt that blockchain technology will change global trade. The question, however, is how it will impact some of the most intractable issues in trade finance. Last year, U.S.$15.5 trillion of merchandise exports were transported around the world. Up to 80% of global commerce requires trade finance to provide liquidity and risk mitigation. However, inefficiencies in trade finance today mean that many applications go unfunded. This U.S.$1.5 trillion trade finance gap is widest in emerging markets and for small- and medium-sized enterprises. Efforts to address these shortfalls have gained limited traction due to the decentralized nature of trade. In this paper, we review the design of enterprise blockchains to explore how changing the architecture of trade finance could impact the drivers of trade finance gaps. By grounding our analysis in the technical architecture of a live, enterprise blockchain platform, we aim to provide a tangible discussion around the technology. Applying blockchain technology to trade finance – regardless of the top of stack application – will directly impact the flow of information, compliance challenges, and profitability in ways that can contribute to a more inclusive trade finance structure.

1. INTRODUCTION

Trade finance reduces risk in the process of trade. Given the number of parties involved, intermediation enables buyers and sellers to transact more efficiently across borders, currencies, and languages. The transaction volumes are huge. In 2017, U.S.$15.5 trillion of merchandise exports were transported around the world across sea, air, rail, and road.[1] Up to 80% of this global trade requires financing.

However, the same characteristics that make trade finance safer also introduce friction and inflexibility. This has resulted in two persistent problems in the sector. First, trade finance is not easily accessible to everyone and in every region. Shortfalls in supply have persistently pooled in frontier markets and among small- and medium-sized enterprises (SMEs). This has direct implications for the ability of emerging markets to capture the benefits of trade driven growth.

Scalability is a second problem. Many believed that digitization was the answer to the lack of visibility, low profit margins, and "know your customer" (KYC) concerns that drive shortfalls. While digitization has changed the way individual entities in trade finance process information, these benefits have not scaled globally into a connected network. If each node in the trade finance network maintains its own proprietary source of information – as it does today – digital documentation needs to be checked and re-entered at every step of the process. Having many different centralized systems globally create localized data centers that do not interoperate with a broader network.

Digital improvements to non-digital infrastructure can only go so far. A fundamental reorganization of the system is required to impactfully change trade finance enough to address the shortfalls and gaps.

Over the past two years, we have witnessed a wholly different solution emerge. Blockchain technology presents an open technology layer that enables programs to connect and scale.[2] The decentralized architecture of a blockchain can serve as a better foundation for interoperation along a global and intermediated process like trade finance.

Trade finance is inherently decentralized; trying to match centralized architecture to this decentralized process has led to the siloes and problems we face today. By changing the structural foundation of trade finance, the technology presents an opportunity to narrow gaps in an unconventional way. Having a decentralized, yet trusted and secure record of information shared between relevant parties can reduce frictions while maintaining the efficiencies of intermediated trade.

In this paper, we take a design approach to explore whether blockchain could rearchitect trade finance to make it more inclusive. This approach is unusual in that we map the reasons for trade finance gaps directly to the features of the technology itself. Our conclusions are thus independent of the specific use case. They apply equally to letters of credit or open account or trade credit insurance. They are applicable in jurisdictions from Brazil to Thailand.

While public blockchains like Bitcoin are the most familiar to the casual reader, we focus in this paper on enterprise blockchains. The reason is that private permissioned blockchains, built with the requirements of companies in mind, are most appropriate for the particular characteristics of trade finance. Trade finance is highly regulated, cross-jurisdictional, and involves multiple parties confidentially exchanging information. In this paper, we will outline advantages that are general to most enterprise blockchain systems, such as IBM's Fabric and Ethereum-based forks,[3] while also mentioning features particular to R3's Corda.[4]

As blockchain technology moves from proof-of-concept to live pilot and beyond in 2018, we can offer insight into whether the technology will ultimately narrow trade finance gaps. We explore the impact of blockchains on three fundamental causes of trade finance gaps: compliance costs, profit, and information. Our objective is to show that the benefits of blockchain technology in trade finance can extend beyond driving operational efficiency to actually narrowing market gaps in frontier markets and among SMEs.

---

[1] WTO, 2017, "World trade in 2016," World Trade Organization, http://bit.ly/2oQZ6IG

[2] All blockchains are distributed ledgers, but not all distributed ledgers "batch" information together into a chain of blocks. For simplicity, the term blockchain and distributed ledger technology (DLT) are used interchangeably in this paper.

[3] An enterprise fork is an adaptation of a public cryptocurrency codebase to make the technology more suitable for companies.

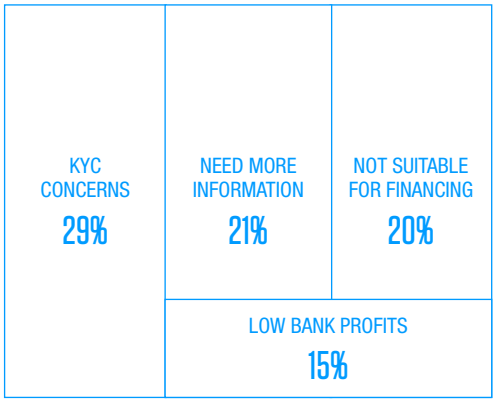
[4] Corda is unique among enterprise blockchains in that it operates a point-to-point transaction model. This means that only participants involved in a given transaction or exchange of data are privy to the data involved in those transaction. In our example, each node transacts on behalf of their clients and shares only the information that is needed to complete the transaction.

2. TODAY'S TRADE FINANCE NETWORK STRUCTURE ENABLES PERSISTENT SHORTFALLS

Trade is conducted through either bank-intermediated risk-mitigating instruments, such as letters of credit, or directly between buyers through open account. While most global trade flows are covered by open account, companies in Asia and the Middle East are heavy users of letters of credit, with 77% of export letters of credit originating in Asia alone. Risk parameters vary depending on when finance and/or risk mitigation is provided, and differ between pre-shipment and post-shipment finance.

Inefficiencies in trade finance means that nearly U.S.$1.5 trillion of demand for trade finance is rejected by banks [ADB (2017)]. The consequence in many cases is that those trades do not happen. A practical example: in a survey of 1,336 firms, respondents report that in 60% of cases when their application for trade finance is rejected, they fail to execute the transaction.[5]

Figure 2: Reasons banks reject trade finance applications (% of rejections)

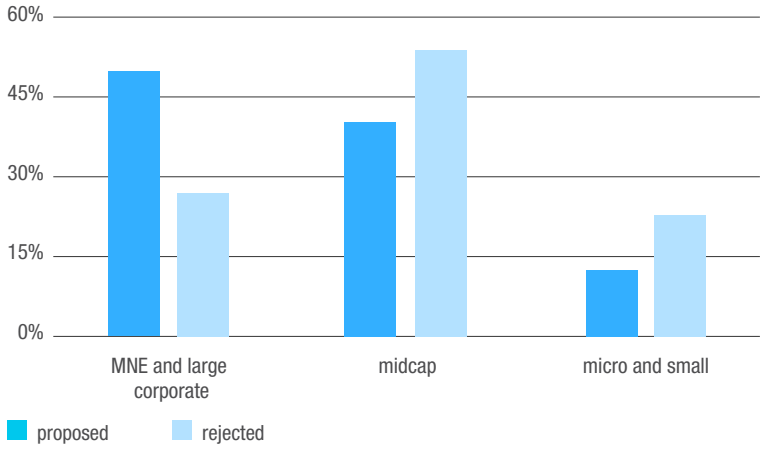


Source: ADB 2017 Trade Finance Gaps, Growth, and Jobs survey

2.1 Trade finance data is centralized to maintain security

In trade finance today, each party to a transaction maintains their own account. These are repeatedly shared, checked for discrepencies, verified, and updated. This process introduces three structural features that contribute to gaps. First, the cost of verifying and checking is high. This is because each individual entity in the transaction needs to ensure that the documents they receive are compliant with regulatory requirements. Entities also need to confirm the information that they expect to see about the transaction – are the goods in the purchase order the same as the goods in the invoice?

Figure 1: Proposed and rejected trade finance transactions (by firm size, 2017)



Source: ADB 2017 Trade Finance Gaps, Growth, and Jobs survey

Second, there is a dependence on correspondent banking limits. Bank-to-bank (correspondent) relationships are central to the current trade finance architecture. Yet thousands of correspondent relations have been severed over the past few years due to cost and regulatory factors. This dependence on correspondent banking networks limits the flexibility of trade finance.

Local or community banks, which are most likely to have banking relationships with SMEs, may not have the necessary correspondent relationships to facilitate an international trade transaction. One global bank reports that the cost of doing due diligence on a bank was of the order of U.S.$75,000 in 2015. When global banks began shedding correspondent relationships in recent years, it was mainly emerging markets that were cut off.[6]

Finally, the use of the cloud, while increasing, is limited in transaction banking. Deploying new digital solutions in banks is slow due to the need to get security approvals. Historically, the standard has been ringfencing data and not allowing any integration. However, given recent high profile hacks, it is clear that centralized data stores are vulnerable regardless of how carefully the data is fenced in.

[5] ADB, 2017 Trade Finance Gaps, Growth, and Jobs survey
[6] IMF, 2016, "The withdrawal of correspondent banking relationships: a case for policy action," International Monetary Fund, http://bit.ly/2abkYln

2.2 Today's architecture is characterized by gaps in emerging markets and SMEs

The way data is shared in trade today exacerbates the inherent challenges of emerging markets. In 2017, the ADB estimated a global trade finance gap of U.S.$1.5 trillion dollars. Furthermore, 40% of global unmet demand for trade finance was pooled in Asia Pacific and Africa.

However, the problem is about more than geography. SMEs in every jurisdiction face shortfalls in access to trade finance. Banks report that 74% of their rejections go to SMEs. These companies worldwide have reported the lack of trade finance as one of the major constraints to their businesses. Globally, they are impacted by the higher cost of screening and higher interest rates. Credit constraints on smaller exporters are higher than those faced by larger firms, to the point of reducing the range of destinations for business or stopping the SME from exporting altogether.

2.3 Digitization improves efficiency, blockchain unlocks trust without centralization

When banks were surveyed on their reasons for rejecting trade finance proposals, their responses fell into three main categories: lack of information, low profit, and KYC concerns (Figure 2). The single undercurrent to all of these causes is lack of visibility into the trade or the client, leading to a "perceived risk" that is higher than what the bank is willing to accept.

Digilization has made important inroads in all of these areas. Regtech seeks to automate KYC reporting, fintech solutions create new sources of information with which to evaluate firms, and digitization initiatives have focused on reducing the cost of trade finance. Each of these has important potential implications for inclusion by disrupting pieces of the trade finance process.

Even as digitization has sought to address parts of the problem, it has also exacerbated its causes in other ways. As banks have implemented new digital solutions, there has been an explosion of destination platforms. That is, the platforms do not interoperate with each other. Digital solutions work as long as all parts of the trade are on the same platform. In global commerce, where a single trade may involve 20 entities, 100 pages of documentation, and 5000 data field interactions,[7] siloed digital solutions make problems worse.

Enterprise blockchains aim to resolve these difficulties with interoperation. There are two features of enterprise blockchains that allow them to retain the benefits of decentralized systems, while addressing the shortcomings of public blockchains for this use case. Enterprise implementations of blockchain technology are better able to meet the data privacy requirements of international trade, as they avoid public broadcast of all information to all parties.[8] Further, depending on the architecture, they can address the scalability limitations of public blockchain systems.[9]

3. DOES BLOCKCHAIN TECHNOLOGY HAVE FEATURES THAT CAN NARROW TRADE FINANCE GAPS?

If today's trade finance architecture enables the persistence of gaps, does this mean that a technology that promised to rearchitect trade finance will narrow them? There is a lot riding on this contention, and some blockchain applications have focused specifically on this area. While we hope that the trade finance applications built on blockchain platforms will accomplish this goal, the sheer scope of different applications makes it difficult to evaluate the potential disruption of each claim.

In this section, instead of looking at the promise of blockchain applications, which is enormous, we look at the mechanics of the technology. Does the technology itself have or enable features that address some of the reasons for trade finance gaps?

By grounding our analysis in the technical architecture of a live, enterprise distributed ledger platform, we aim to provide a tangible discussion around the technology of blockchain. Some of these benefits are inherently enabled by the blockchain platform itself, while others allow producers of applications to drive the benefits. We cover the three drivers from Figure 2: low profitability, regulatory concerns, and information.

---

[7] BCG, 2017, "Digital innovation in trade finance: have we reached a tipping point?" Boston Consultancy Group, SWIFT focus whitepaper.

[8] No risk department would agree to having each node contain identical copies of the entire transaction history. Even though there is some technology being introduced in public blockchain systems to address privacy concerns, the technology can be too immature for enterprises. In addition, often enterprises do not want information shared with other parties, even if that information is "encrypted." Corda technical Whitepaper, http://bit.ly/2txmifg9

[9] Buterin, V., 2016, "Ethereum platform review," R3 Research Papers

3.1. Low profitability

Low profitability comes from two sources: the bank's cost of processing a transaction and the expected revenue. SMEs will naturally provide a smaller expected profit as their transaction volume and frequency may be expected to be low.

The cost and time to process a trade transaction can be significant due to the variety of actors and steps that are involved. Much of the cost arises from delays, friction, and additional effort needed to handle trade data. This is underpinned by the fact that trade finance is a linear process that is heavily reliant on paper documents. The paper documentation is carried from port to port along with the cargo, checked, signed, faxed to the various parties, including banks, with very little visibility of the whole process by any single participant. Manual checking is time consuming, and can be error prone. Bottlenecks also occur frequently because no party has overall control over or visibility into the full process.

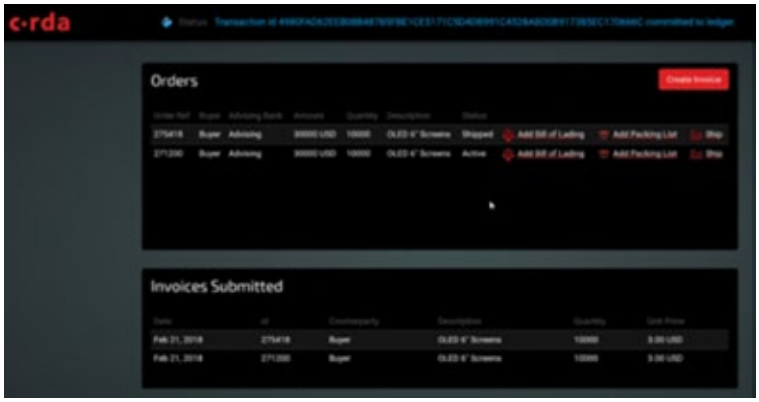
Delays can have real cost implications. Take demurrage, a payable charge to the owner of a chartered ship in respect of failure to load or discharge the ship within the agreed time. These charges can add up to U.S.$150 per container per day,[10] which may not seem like much, but when applied to a whole vessel, such as a Panamax fleet ship, the cost can reach U.S.$750,000 per day of delay.

These costs are further exacerbated by distant markets, banks that lack direct correspondent relationships, and small firms. For these entities, the time needed to execute a transaction, the prevalence of errors and amendments, and the need to verify all parts of the transaction due to the need to establish provenance of paperwork[11] contribute to low profitability.

There are several technical elements built into enterprise blockchains that impact the cost and time of processing. Steps are eliminated when relevant parties maintain the same version of the truth. Having automatable processes will speed up steps that cannot be eliminated. Both oracles and signatures (or multiple signatures) from trusted parties can trigger other events on a blockchain, removing delays for information dependencies.

Blockchain features that can address profitability are: (1) single truth layer, (2) automatable processes, (3) oracles, and (4) signatures, multiple signatures (multisig).

Figure 3: Status of different orders shared with relevant parties



Source: Letter of credit demo, http://bit.ly/2FVeDz9

All of these features lead to the elimination of confirmation steps as parties can trust that the information that they see is the same as other parties.

One of the major cost multipliers in trade transactions is the passing back and forth of the same documents multiple times. For example, there are typically 19 steps (without amendment) in a letter of credit. By having a single truth layer, we can potentially eliminate seven of these steps immediately.[12]

As one example, Figure 3 illustrates the seller's screen in an on-ledger letter of credit transactions. This is a snapshot of the key data forms that are included.

In any transaction on the blockchain, there is a single source of verified data that is immutable. As a result, all parties can have confidence that the information on the screen is verified, and is the same as what their counterparties see on their screens. The data is accurate and reliable from the beginning.

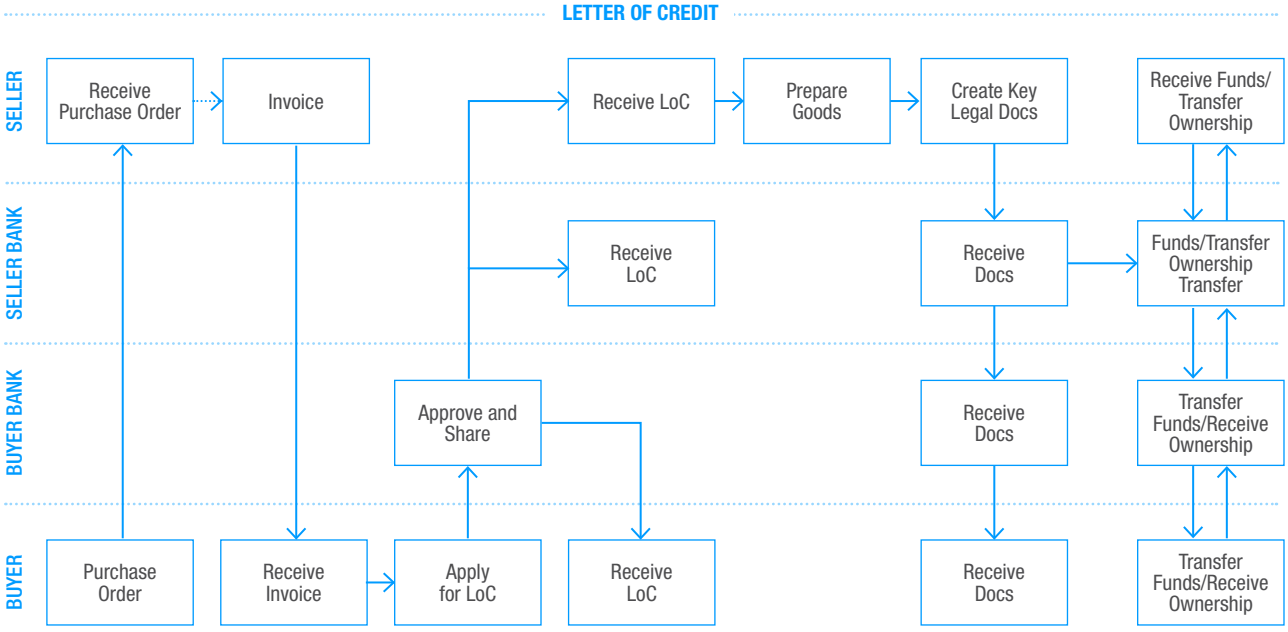
Having a single source of truth shared between relevant parties reduces costs that are due to errors associated with repetitive verification. This will shorten the time to financing because it reduces information float (the time until information is available) and increases real time visibility of trade. Faster information transfer with fewer errors will reduce the additional costs associated with riskier trades.

[10] Czajkowski, A., "Demurrage, detention, per diem… Oh my! 6 tips to avoid additional charges," Shapiro, http://bit.ly/2H5WEFR

[11] Default is not the reason that cost and profit is a problem. Default rates for trade finance are below 1%, and recovery even then is on the order of 80% (ICC trade register). The problem of cost and profitability is related to paperwork and verification and checking.

[12] Author's calculations using a pilot version of a Letter of Credit on Corda.

Figure 4: Flows between nodes on Corda along trade finance lifecycle



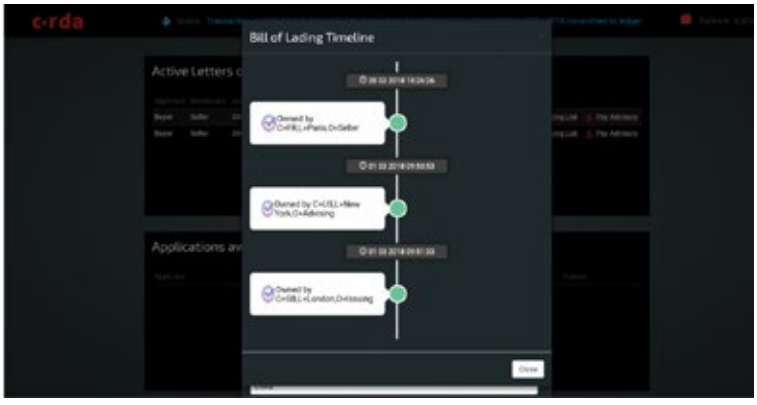
Source: R3

A second way that blockchain can address profitability is by automating some steps that currently have a long lag time. On Corda, flows enable coordination from different nodes to trigger "transactions," or updates to shared states between parties that are automated. Figure 4 illustrates the flows in a letter of credit trade lifecycle.

Oracles and third party signatures are two examples of inputs to a blockchain that can trigger an automated process. An oracle is an agent on the blockchain that provides information to the participants of one or more business networks. They source information from real world events, third party data providers, or other blockchain activities. An oracle can push information to a business network either regularly or upon request, and is not party to a trade.

Oracles are meant to provide facts to business network participants. Figure 4 shows an example where an electronic bill of lading (eBL) provider acts as an oracle in a trade transaction. In a trade transaction, ownership of the goods changes throughout the process. The document of title – in this case represented by an eBL – allows the bearer to take ownership. After the goods are ready to ship, the shipper requests the carrier to draft a bill of lading. The shipper exchanges the goods for the eBL from the carrier. The shipper then presents the eBL

Figure 5: Bill of lading timeline with title transfer (illustration of DLT LC platform)



Source: Letter of credit demo, http://bit.ly/2FVeDz9
Notes: O=owner

to the advising bank in fulfillment of the documentary submission requirements of a letter of credit. Since each eBL is associated with a unique title registry record – which is maintained by the oracle – the transfer of title from the shipper to the advising bank only occurs after querying the oracle to get confirmation of the registry record number. This also introduces additional visibility into who owns the goods throughout the process.

A trusted third party may also provide signatures to a blockchain. For example, DHL may send a pay out upon signature when goods have arrived at a certain port, triggering the next process automatically, removing frictions. Once the physical cargo has been checked and the data input into the system, an automated contract could instantaneously release the funds from the seller's bank to the buyer's bank. This is one example of a smart contract, and a self-executing transaction. This could significantly reduce the delay between the checking of the cargo and the final release of funds.

3.2 Lack of information (risk management)

A second driver of trade finance rejections is a lack of reliable information. This makes it difficult for stakeholders to accurately measure risk, a problem that is compounded for SMEs.

Where risk is the result of information asymmetries, blockchain can improve risk management. It could enable us to answer questions like: Can this SME perform. Is it capable of delivering quality goods/service in the timeframe stipulated in a contract. Will it remain solvent for the duration of its obligations?

Recording transactions on a blockchain leads to a treasure of metadata from which financial institutions could answer SME performance risk related questions reliably and efficiently. If all transactions in an SMEs life are captured step by step through a blockchain, that information could be organized to address a root cause of SME finance market gaps.

> "Trade finance is inherently decentralized; trying to match centralized architecture to this decentralized process has led to the siloes and problems we face today."

In addition, the ability to identify the legal entity of a participant in a transaction is fundamental to efficient trade finance. It can act as an organizational anchor – or a master key – from which all other information can be hung. Corda uses X.500 distinguished names to identify participants. To further improve transparency about network participants, many solutions then link to "legal entity identifiers" (LEI).

An LEI is a 20 digit, alpha numeric code. It is connected to key reference information that allows for the unique identification of legal entities participating in financial transactions. The LEI verifies, on an annual basis (i) who's who; (ii) who owns whom; (iii) who owns what. The body responsible for administering the LEI system is, itself, regulated by over 70 central banks around the world.

An important benefit of being able to organize data around entities with canonical identifiers that are used universally across the globe, is that it becomes possible to develop a map that charts out the history of all transactions performed across all participants. This becomes a very powerful tool for reducing fraud, because all the parties involved in a trade are known and have been validated.

3.3 Compliance (KYC concerns)

Regulatory oversight plays a critical role in the functioning of the global financial system. Over time, both the diversity of regulations and the size of sanctions fines have increased.

This relates to the third major driver of rejections for trade finance proposals – KYC and money laundering concerns. The cost and complexity of regulatory compliance play an important role in transactions costs. The problem for trade finance is that 29% of rejections are based on KYC concerns.

Enterprise blockchains can address the uncertainty related to compliance via three features. These include live information sharing through a regulator node, active regulation by requiring attestations by third parties that have done KYC checks, and more data for retrospective analysis that may facilitate more effective regulation.

Beyond blockchain's ability to enhance the realiability and efficiency of conducting KYC and anti-money laundering due diligence, the technology can also address two related regulatory issues. The first is uncertainty by regulators. Because regulators only see a trade after it has occurred, their lack of insight shifts the burden of vigilance onto banks. The second is uncertainty by banks. The regulatory environment is complex and not harmonized. Each bank needs to satisfy different levels of regulation.[13]

Blockchain features that can address regulatory compliance are: (1) notary functionality, (2) regulator nodes, (3) attestations, and (4) audit trail.

3.3.1 NOTARY AGREEMENT PROTECTS AGAINST DOUBLE-INVOICING OF THE SAME STATE

Double-invoicing is a common concern in trade finance. It can occur in error where a transaction is mistakenly counted twice, or in a fraudulent setting where a malicious actor intends to game the process. This requires financial institutions to spend time to validate all the transactions to prevent instances of double-invoicing. Standard Chartered, for example, lost almost U.S.$200 million from a fraud involving counterfeited paperwork, where different banks and trading houses were holding separate titles for the same metal at China's Qingdao port in 2014.

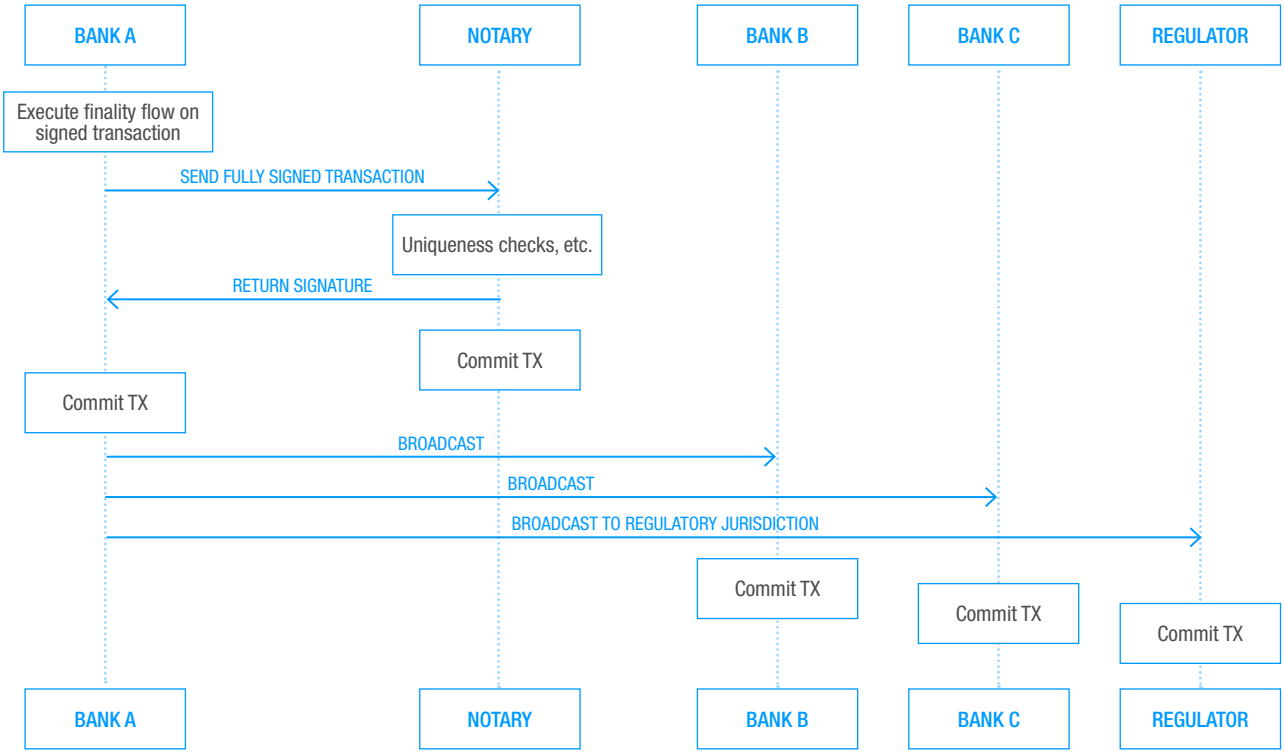
Blockchain technology has the potential to address some of these fraudulent transactions. With Corda, a notary ensures uniqueness of an invoice or payment. This has the potential to make important inroads on the double-spending problem. Because a transaction is represented as a particular state, it is mathematically impossible to re-use the same state more than once. That is, if a particular invoice has a certain number, the same asset literally cannot be sent to two banks – it can only be used once, and the notaries will enforce this. While this does not protect against people creating two separate invoices with different numbers, banks can ensure that a certain state is only used once if they are on the same application.

3.3.2 LIVE REGULATOR OVERSIGHT OF PARTICULAR IDENTITIES OR TRANSACTIONS

As regulations around trade finance continue to grow, the sanctions for violations have also grown. According to FinCen data, the number of suspicious activity reports grew exponentially from 1990 to 2013. In addition, market participants have to generate compliance reports on a regular basis for submission to the relevant regulators. Much of this work is still manual, which leads to high overhead costs.

[13] In a recent informal survey of global banks, repondents each had a different process for KYC in a supply chain finance transaction.

Figure 6: Corda flow with a regulator node



Source: R3

Enterprise level blockchains enable the addition of a specific type of entity into the transaction that we refer to as a regulator node. This regulator node can be incorporated into the network, and it enables regulators to monitor transactions that occur on a real-time basis, but does not give them the ability to change the transaction. This can reduce the need for manual regulatory reporting and can significantly reduce costs.

Blockchain addresses the cost and complexity of regulatory compliance in an unusual way. While existing regtech solutions have focused on simplifying KYC, this is only one part of a complex problem. By allowing regulators to have direct insight into transactions, the reporting process can become less burdensome by becoming incorporated into the transaction itself.

3.3.3 SIGNATURES BY THIRD PARTIES THAT HAVE PERFORMED DUE DILIGENCE

On a blockchain network, trusted third parties can attest that a particular party is not nefarious. Onboarding of a new corporation, or entity, could involve the signature of an entity that has done due diligence.

3.3.4 MORE DATA CAN ALLOW BETTER RETROSPECTIVE ANALYSIS TO MAKE FUTURE PROJECTIONS OF CRIMINAL ACTIVITY

Immutability of data refers to the fact that a state cannot be changed or modified after it has been created. This creates a clear audit trail into the transaction, as historic states are stored and can be accessed. It will be impossible to tamper with data, and data integrity can be maintained at all times.

4. CONCLUSION

Trade finance gaps will not be resolved until trade finance changes. This kind of thinking is not new. But the technology to make it happen is. Blockchain will impact how trade finance is done. It will become safer, faster, and more secure as banks and corporates move trade onto the blockchain.

Does this mean that blockchain will solve the trade finance gap? Not alone. As it spreads, its hard coded features will improve the potential for trade finance to be more inclusive and available. To take full advantage of the benefits that blockchain has to offer in trade finance, we need to consider three questions while designing top of stack applications: Can outside providers easily contribute data? Does it follow or reuse existing standards and contracts? Will it improve information and data flow?

2018 is the year that proof-of-concepts are moving into pilots and production. As we have learned during the design phase, the problems in trade finance are going to be much harder to solve than they look on the surface. As trade continues to evolve, it has become increasingly urgent that financial institutions are equipped to deliver not only the types of financing needed today, but also the types of financing that will be needed tomorrow.

The aftermath of money market fund reform

JAKOB WILHELMUS | Associate Director, International Finance and Macroeconomics, Milken Institute
JONATHON ADAMS-KANE | Research Economist, International Finance and Macroeconomics, Milken Institute[1]

ABSTRACT

Extensive regulatory overhaul changed the money market fund (MMF) industry considerably, especially for institutional clients. Nonetheless, MMFs continue to be an important cash management tool for institutions even though their asset allocations are now much more restricted to preserve the feature of a constant share price. This article shows how regulatory changes to MMFs correctly remove unviable promises of immediate liquidity at a constant share price while holding asset portfolios with varying risk exposures. We emphasize the importance of allowing price signals to reveal the impact of changes in the risk environment on asset holdings. We also believe that quantitative restrictions (e.g., withdrawal fees and gates) are counterproductive for preventing runs – they do not aid price discovery, and incentivize investors to circumvent the restrictions to access their otherwise liquid assets in times of heightened liquidity demand. These shifts in the money market and related channels of short-term financing should act as a reminder that regulatory pressure on one part of financial markets has repercussions throughout the financial system, leading to unexpected adaptation by market participants.

[1] This article is an updated version of the Milken Institute report titled "Regulation almost destroyed money market funds, but cash management needs kept them alive."

1. INTRODUCTION

Extensive regulatory overhaul in October 2016 changed the money market fund (MMF) industry considerably, especially for institutional clients. Nonetheless, MMFs continue to be an important cash management tool for institutions even though their asset allocations are now much more restricted to preserve the feature of a constant share price.

Key regulatory changes were threefold. First, institutional prime MMFs must float their net asset value, abandoning their signature feature of a constant share price. Second, institutional prime MMFs must adopt a system of redemption gates and fees to ensure sufficient liquidity. Third, government and retail MMFs are exempt from the floating NAV requirement and from redemption fees and gates.

Following the October 2016 reforms, institutional investors made significant changes to their MMF investments. They faced a choice of shifting their investments to government MMFs (offering a stable share price) or remaining invested in higher yielding prime funds (now with a floating share price). Institutional depositors overwhelmingly favored retaining a constant share price even if returns were lower: institutional prime funds lost almost 74% of their net assets to government funds and partly to retail prime funds. This reallocation shows that immediate liquidity at par dominates slightly higher returns when it comes to the needs of institutional investors' cash management.

As we show below, regulatory changes to MMFs correctly remove unviable promises of immediate liquidity at a constant share price while holding asset portfolios with varying risk exposures. We emphasize the importance of allowing price signals to reveal the impact of changes in the risk environment on asset holdings. We also believe that quantitative restrictions (e.g., withdrawal fees and gates) are counterproductive for preventing runs: they do not aid price discovery, and incentivize investors to circumvent the restrictions to access their otherwise liquid assets in times of heightened liquidity demand. More specifically:

- New MMF regulations acknowledge that shares in prime MMFs are subject to both market and credit risk. The rise in the rates offered by non-government MMFs helped stem the outflow of assets to government MMFs. At the same time, demand for U.S. Treasury bills (and agency debt) that removed credit risk from government MMF portfolios increased greatly.
- More concerning is the impact of liquidity constraints, through fees and gates, and the prospect of extending them to mutual funds in general. These non-price mechanisms are designed to limit investors' access to their assets, particularly during periods of market turmoil.

In the remainder of the paper, we describe the asset shifting by MMFs as well as its impact on the different markets. Section 2 outlines how the approximately U.S.$1 trillion that shifted from prime to government MMFs have affected commercial paper and deposits. We then provide an overview of the asset reallocation into government funds, before concluding.

2. GOVERNMENT MMFS DISPLACED PRIME MMFS AND ALLOWED INSTITUTIONAL CASH MANAGEMENT TO RETAIN REDEMPTIONS AT PAR

Stability is a key characteristic for cash management tools. Previously, MMFs provided stability by maintaining a constant share price as long as mark-to-market net asset values, rounded to the nearest one percent, would yield the same price – a key exemption authorized under rule 2a-7.[2]

Share price stability offered by prime MMFs conveyed a false sense that MMF shares are a risk-free asset. However, prime MMFs held portfolios that can change so dramatically in value that the dollar parity under 2a-7 cannot hold. Before the reforms, corporate treasurers chose to deposit most of their funds into higher-yielding prime funds over more prudent government funds because both promised redemption at par without restrictions.[3]

These shortcomings became unsustainable during the financial crisis in 2008 when some prime funds were no longer able to maintain a constant share price. The Securities and Exchange Commission (SEC) adopted amendments to reduce the risks of MMF runs that could cascade into a mass sectoral asset reallocation with systemic consequences.[4] These new regulations stripped away the constant share price characteristics

[2] MMFs had to constantly calculate a shadow price using available market prices or fair value pricing.
[3] Prime MMFs primarily invest in corporate debt whereas government MMFs invest in government and agency debt (or repos of the respective securities).
[4] Rule 2a-7 Amendments by SEC in July 2014.

of institutional prime MMFs and imposed redemption gates and fees. Institutional depositors reacted by shifting almost exclusively to government MMFs to preserve redemption capabilities at a constant share price without other restrictions. Although the change in regulation was expected to cause a reallocation from prime to government funds, the magnitude of the change has caught many by surprise.[5] Approximately U.S.$1 trillion shifted from prime to government MMFs (Figure 1).[6]

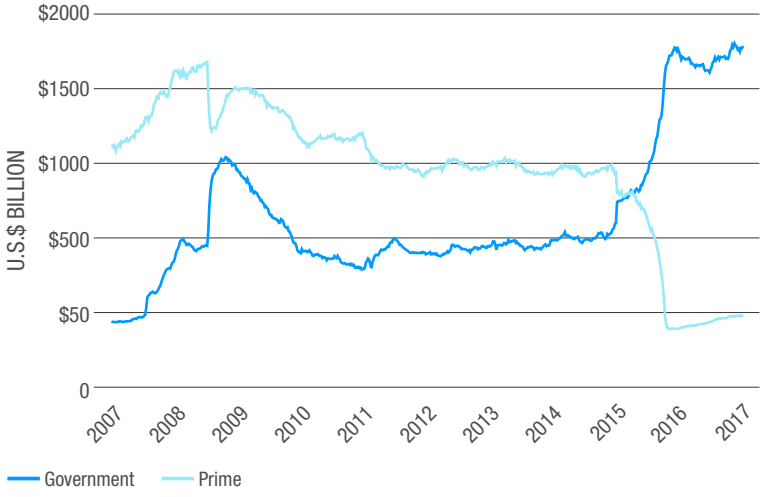
3. MMF INVESTMENTS CHANGED SHORT-TERM FUNDING OPTIONS FOR BANKS

3.1 Prime funds – the drawdown

The reallocation of U.S.$1 trillion from prime to government MMFs had a substantial impact on market demand for the underlying instruments. New roles of MMFs consequently changed the mix of instruments by which borrowers raised short-term funds. MMFs hold a variety of short-term instruments – government issued and backed securities, commercial paper, certificates of deposits, and repurchase agreements. Most prime funds invest largely in higher-yielding commercial paper (CP) and certificates of deposit (comprising around 60% of their total assets). From the issuer's perspective, almost 40% of total CP was held by MMFs. However, following the MMF reforms, this share has fallen to 13%, or U.S.$210 billion as of December 2017 (Figure 2).

Most CP is issued by banks – and this accounted for most of the decline in MMFs' holdings following the reforms (Figure 2).[7] Foreign banks' ability to raise short-term funding was handicapped more than

Figure 1: MMF – regulatory impact



Source: Federal Reserve

domestic banks. This is because domestic banks had alternative funding sources, such as advances from the Federal Home Loan Banks (FHLBs) and had already been gradually switching funding sources away from issuing CP for reasons unrelated to the 2016 reforms. FHLB advances became available at a lower price and were extended for terms (lengths of time), which proved useful for meeting liquidity requirements under Basel III.[8] In contrast, foreign banks are not able to access

[5] Remarks by S. Potter, Executive Vice President of the Markets Group of the Federal Reserve Bank of New York, at UCLA, April 2017.

[6] The word "shift" should not be taken to mean a one-to-one movement of investment in prime funds to government funds, as such information is not available.

[7] Banks are generally prohibited from issuing CP themselves, but can raise funds through asset-backed CP issued by conduits, or financial CP issued by bank-related finance companies held by the parent bank holding company [Kacperczyk and Schnabl (2010)].

[8] Federal Housing Agency Office of Inspector General (2014).

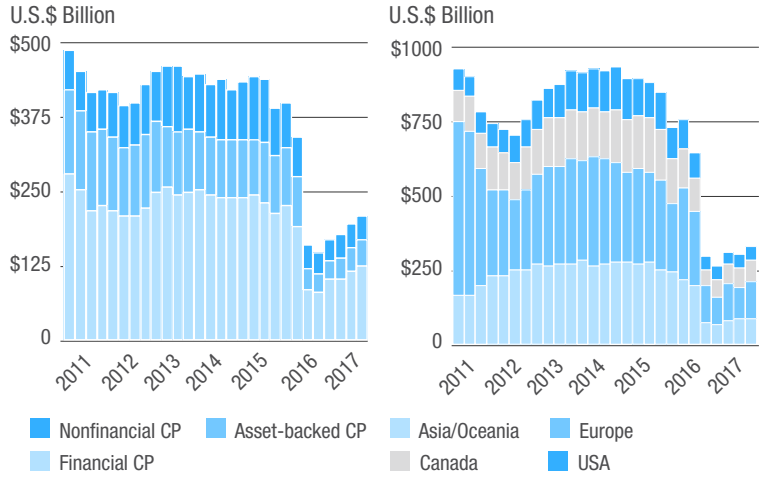
FHLB advances and, therefore, had no alternative way to raise short-term funding other than through their CP issuance. Consequently, as prime funds withdrew from the CP market and also reduced their deposits (Figure 2), the reserves and overall balance sheets of foreign banks' U.S. branches contracted.

3.2 GOVERNMENT FUNDS – ASSET REALLOCATION AND FHLBS

The bulk of outflows from prime funds went into government funds, which accommodated the inflows by increasing purchases of Agency and Treasury debt, and using repurchase agreements through the Federal Reserve's overnight reverse repo facility. As government money market funds' portfolios grew on aggregate, the proportion of their investment allocated to agency debt and agency-backed repos stayed persistently high, accounting for 44% of their assets as of October, 2017 (Figure 3).

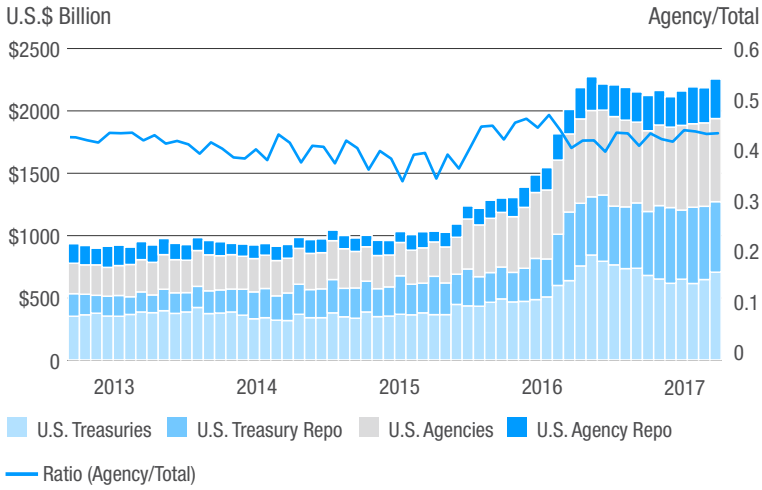
As MMFs' demand for agency debt grew and their demand for CP fell, domestic banks adjusted their funding structures accordingly. Banks increased their borrowings – called advances – from FHLBs, as a ready substitute for raising funds by issuing CP. FHLBs' issuance has increased, particularly their short-term, floating rate obligations, which are eligible to MMFs – outstanding floaters increased from U.S.$80 billion at the end of 2015 to U.S.$295 billion by June, 2017.[9]

Figure 2: Prime MMFs' CP holdings (left) and deposits by domicile (right)



Source: SEC and Federal Reserve

Figure 3: Government MMFs holdings



Source: Federal Reserve

[9] FHLBanks Office of Finance Monthly Issuance Data Reports. "Short-term," here, refers to 397 days or less to maturity.

4. CONCLUSION

Price stability is an essential characteristic of a cash management tool. However, price stability may induce investor complacency by introducing the incorrect notion that underlying assets held by a MMF are risk-free. This distortion can induce destabilizing runs in times of extreme financial stress. By allowing the share price of MMFs to vary, new regulations have highlighted the fact that shares in prime MMFs are not risk-free. This change in regulation led to a U.S.$1 trillion reallocation from prime to government funds, thereby reducing the risk of runs caused by the false sense of security of a guaranteed fixed share price when market conditions become volatile.

Fees and gates, the second pillar of the new MMF regulations, may stem runs temporarily. However, they may induce attempts to circumvent the restrictions and could make a liquidity crunch worse by cutting off investors from accessing their liquid assets just when liquidity is scarce. Only institutional prime MMFs remain subject to the rules on gates and fees. However, regulators may extend these quantitative restrictions on withdrawals to mutual funds more broadly. Such a regulatory shift might create preemptive runs, as the option to suspend convertibility introduces potential restrictions on investors' access to their assets in times of stress. In other words, investors might withdraw their investments if the likelihood of redemption restrictions increases substantially. The almost-disappearance of institutional prime funds is an indication of the importance investors place on having reliable access to their assets.

These shifts in the money market and related channels of short-term financing should act as a reminder that regulatory pressure on one part of financial markets has repercussions throughout the financial system – leading to unexpected adaptation by market participants. To cite Fed vice chair Fischer, "[w]hile the current configuration of money markets reveals a reduced financial stability risk […] this configuration may not yet represent the final equilibrium."[10]

[10] Stanley Fisher (2017).

References

Aldasoro, I., T. Ehlers, E. Eren, and R. McCauley, 2017, "Non-US banks' global dollar funding grows despite US money market reform," BIS Quarterly Review, March, 22-23

Bank for International Settlements, 2017, BIS Quarterly Review, June

Craig, B., S. Millington, and J. Zito, 2015, "Who is holding all the excess reserves?" Economic Trends, Federal Reserve Bank of Cleveland, August

Federal Housing Finance Agency Office of Inspector General, 2014, "Recent trends in federal home loan bank advances to JPMorgan Chase and other large banks," Evaluation report EVL-2014-006, Federal Housing Finance Agency

Fischer, S., 2017, "An assessment of financial stability in the United States," IMF Workshop on Financial Surveillance and Communication, June

Kacperczyk, M., and P. Schnabl, 2010, "When safe proved risky: commercial paper during the financial crisis of 2007–2009," Journal of Economic Perspectives 24:1, 29-50

Kreicher, L. L., R. N. McCauley, and P. McGuire, 2013, "The 2011 FDIC assessment on banks' managed liabilities: interest rate and balance-sheet responses," BIS working papers 413, Bank for International Settlements

McCauley, R. N., and P. McGuire, 2014, "Non-U.S. banks' claims on the Federal Reserve," BIS Quarterly Review, March, 89-97

Securities and Exchange Commission, 2012, "Response to questions posed by commissioners Aguilar, Paredes, and Gallagher," Division of Risk, Strategy, and Financial Innovation, U.S. Securities and Exchange Commission

Stigum, M., and A. Crescenzi, 2007, Stigum's money market, McGraw-Hill, Fourth Edition

Costs and benefits of building faster payment systems: the U.K. experience[1]

CLAIRE GREENE | Payments Risk Expert, Federal Reserve Bank of Atlanta
MARC RYSMAN | Professor of Economics, Boston University
SCOTT SCHUH | Associate Professor of Economics, West Virginia University
OZ SHY | Author, How to price: a guide to pricing techniques and yield management

ABSTRACT

A number of countries have implemented faster payment services that allow consumers and businesses to rapidly transfer money between bank accounts. These services compete with slower, existing payment services. In 2008, the U.K. implemented its Faster Payments Service (FPS) at a cost of less than £200 million (.014% of U.K. GDP, or U.S.$307 million) spread over seven years, plus investment costs borne by each participating bank to connect to the FPS. This paper examines the economic cost-benefit analysis underlying the U.K. FPS investment decision and describes the subsequent diffusion and use of FPS through 2014.

[1] This paper was written while the authors were members of the Consumer Payments Research Center of the Federal Reserve Bank of Boston. We thank William Murdock III, Christine Marieni, and Michael Corbett for excellent research assistance, Suzanne Lorant for excellent editorial services, and Jim Cunha and Bob Triest for their comments and suggestions. We also thank Jim Mortimer, Head of International Propositions at VocaLink, Kris Kubiena, Proposition Delivery Director at VocaLink, Alex Smith, Marketing Manager at the Payments Council, Mike Banyard, Head of Development, Faster Payments Scheme, and Gordon Baird, Chief Executive Officer at Independence Bancshares, Inc.

An expanded version of this paper is available on the website of the Federal Reserve Bank of Boston at https://www.bostonfed.org/publications/current-policy-perspectives/2014/costs-and-benefits-of-building-faster-payment-systems-the-uk-experience-and-implications-for-the-united-states.aspx.

1. INTRODUCTION

A number of countries have implemented faster payment services that allow consumers and businesses to rapidly transfer money between bank accounts, in transactions known as "account-to-account (A2A)" payments.[2] These services can be provided by banks or nonbanks (with cooperation from banks) that are connected to a new central infrastructure[3], which supports faster authorization, clearing, and settlement (ACS) than existing payment services, along with faster confirmation to the payer and payee of each money transfer.

In 2008, the U.K. adopted and implemented a new Faster Payment Service (FPS) rather than investing in improvements to speed up its existing payment system.[4] According to VocaLink (2009), a key motivation for the U.K. decision was a request by the Office of Fair Trading (OFT) to remove the float from standing orders (regular recurring payments for a set amount) in the U.K. banking industry. It is not known whether British banks or nonbank payment service providers would have taken this step eventually without the directive of government authorities.

This paper examines the economic cost-benefit analysis underlying the U.K. FPS investment decision.[5] We report quantitative estimates of the monetary costs associated with the FPS but only qualitative descriptions of potential benefits to all parties involved: consumers, merchants, financial institutions, other businesses, and government. This analysis provides a framework that may help the payments industry and payments policymakers to assess the viability, costs, and benefits, and social welfare of adopting faster payment services.

The cost to U.K. banks of building, installing, and maintaining the British FPS was relatively modest. According to sources at VocaLink, which operates the infrastructure of the U.K. FPS, it cost less than £200 million (U.S.$307 million), or .014% of U.K. GDP, to install and operate the FPS for the initial contract period of seven years (2008–2015), plus estimated investment costs of up to £50 million (U.S.$77 million) for each participating bank to connect to the FPS.[6] Thus, the estimated maximum total cost of FPS was less than .06% of U.K. GDP in 2008. For payment system participants, introduction of the new FPS may also have led to revenue transfers to the FPS or losses associated with substitution from existing payment methods, but the U.K. data suggest that revenue effects from substitution have been relatively small thus far. Direct revenue from the use of FPS during the initial period was zero, because users do not pay for each transaction. Costs and revenues beyond 2015 are not known at this time.

Identifying potential benefits from faster payments is more challenging, and currently it is impossible to produce complete, accurate, and precise quantitative estimates.[7] Instead, this paper uses survey data [Faster Payments (2013)][8] on the use of FPS by U.K. participants to describe its diffusion through the end of 2013. Then it surmises what types of benefits may ensue to whom when A2A payments are made faster from authorization to settlement or when confirmation of payment is communicated faster.[9]

"A new payment technology like FPS may yield additional benefits beyond speeding up individual payments."

A new payment technology like FPS may yield additional benefits beyond speeding up individual payments. While these benefits are even harder to assess and measure, they may be more important than speed per se. The following apocryphal quotation, attributed to auto maker Henry Ford, illustrates the difficulty in assessing benefits of products and services before consumers can actually experience them: "If I had asked people what they wanted, they would have said faster horses." Nevertheless, faster A2A payments could provide the following benefits: (1) facilitate business-to-business (B2B) payments; (2) facilitate mobile payments, a rapidly developing payments application; (3) improve payment

[2] These countries include Singapore, Mexico, India, South Africa, and Switzerland, as well as the U.K. A detailed description of the last four systems is given in Summers and Wells (2011) and Jacob and Wells (2011). Faster payments systems vary in their functionality and use. Lodge (2014) identifies more than 35 faster payments systems around the world. Clear2Pay (2014) cites about a dozen systems.

[3] "Infrastructure" refers to the servers, software, and communication networks that connect participating financial institutions and transmit payment messages from the sending account to the receiving account and back.

[4] For detailed evaluations of the British FPS, see Milne and Tang (2005) and Summers and Wells (2011).

[5] A similar study was done by the Reserve Bank of Australia; see RBA (2012, p. 1), which states: "While not wishing to dictate how the strategic objectives are met, the paper also outlines the Board's thinking on a possible approach to architecture for providing real-time payments."

[6] All values in British pounds are converted to U.S. dollars using the OECD's estimates of the PPP dollar-pound exchange rate (see http://bit.ly/2CWQt4b).

[7] Stavins (1997) describes a similar challenge in examining the costs and benefits of switching from paper check presentment to electronic check presentment with truncation.

[8] 2012 data is used because the Faster Payments Tracking Survey was discontinued in later years.

[9] The value of speed may be different for recurring bill payments than for payments made at the point of sale (POS). A full analysis of this differential is beyond the scope of this paper and little research is available for bill payments. For more detailed research on POS speed, see Klee (2008), Borzekowski and Kiser (2008), Schuh and Stavins (2015), and Polasik et al. (2013).

security;[10] (4) be available at all times (24/7/365); (5) facilitate person-to-person (P2P) payments, which are typically handled by cash and check in the U.S.; and (6) facilitate faster international payments using standards such as ISO 20022.

Although we cannot provide monetary estimates of the benefits of the U.K. FPS, the total costs of the new system relative to the U.K. population (63 million in 2012) suggests that the value of benefits per individual per year need not be large to give FPS a positive net present value. Because the U.K. payment system prior to the FPS bears striking similarity to the current U.S. payment system, the U.K. experience has implications for the U.S. payments industry.[11]

2. CONCEPTS AND DEFINITIONS

2.1 Payment system

According to the Bank for International Settlements [BIS (2003)], a payment system consists of a set of instruments, banking procedures, and, typically, interbank funds transfer systems that ensure the circulation of money. Summers (2012) uses a much broader definition, where a payment system is an infrastructure (consisting of institutions, instruments, rules, procedures, standards, and technical means) established to effect the transfer of monetary value between parties who are discharging mutual obligations.

Each payment activity is divided into steps. Different payment instruments may use different steps to accomplish money transfers originated by a payer and received by a payee. Moreover, even if two payment instruments use similar steps, they may vary significantly with respect to the time it takes to accomplish each step. The definition of steps also depends on the role played by the entity participating in the payment activity.

2.2 Authorization, clearing, settlement, and notification

Electronic transactions are generally divided into three major steps: authorization, clearing, and settlement. Table 1 presents a possible timeline for an FPS transaction and compares it with a typical debit card transaction.

Two important points need to be made regarding Table 1. First, the term "clearing" is used differently in discussions concerning the British FPS and debit card transactions. In the U.K. FPS process, "clearing" occurs when end users are debited and credited. In debit card transactions, clearing refers to the exchange of data between the card issuer and the card acquirer. Second, in the U.K. FPS process, the payer and the payee are debited and credited before banks settle their funds transfers. This need not be the case for debit cards.

Three key terms characterize electronic funds transfers.[12]

Table 1: Possible timelines of U.K. FPS and debit card transactions.

STEP	FPS (U.K.)	DEBIT CARD (SINGLE MESSAGE)
1	**Request:** payer submits payee's bank account details and amount.	**Authorization (approved/declined):** card swiped at POS, typed online, or provided over the phone. Issuing bank may put a $1 to full amount hold on payee's account. **Clearing:** data exchanged provide the verification for the dollars debited from issuing banks and credited to acquiring banks.
2	**Clearing or rejection:** funds withheld from payer's account and credited to payee's account.	**Settlement:** aggregated netted funds transfers among banks. Transfers include interchange fees from the acquirer to the issuer.
3	(Possible time gap until settlement.)	(Possible time gap until settlement ends. Few hours or longer.)
4	**Settlement:** funds transfers among banks (three times daily during business days).	Within 24 hours, funds released by acquirer or card processor are credited to the payee's account.

Sources: VocaLink (2009) and Herbst-Murphy (2013), mostly pp. 12–14.

[10] A detailed discussion of security is outside the scope of this paper.
[11] In 2012, the Federal Reserve announced an updated strategic plan that emphasized a preference for faster U.S. A2A payments. More recently, the Fed set forth a "vision to improve the speed and efficiency of the U.S. payment system from end to end" [FRS (2013, p. 2)].
[12] The definitions of clearing and settlement are taken from BIS (2003). BIS (2003) does not define authorization; hence, the reader is referred to http://bit.ly/2l3Cqh6. In the context of debit card transactions, Herbst-Murphy (2013, p. 1) refers to authorization as the creation of electronic records in the merchant's transaction system and at the cardholder's bank.

Authorization: "Giving power or permission to (someone or something)." At the POS, authorization begins when the payer swipes a card. For electronic A2A transfers, a payer (fund sender) may use online, ATM, phone, or a mobile device to fill out a form instructing a financial institution to transfer funds. The payer generally has to click (or press) on a "confirm" button, thereby having a second chance to cancel the authorization. Authorization for online debit card transactions is similar; however, at the POS, authorization begins when a card is swiped.

Clearing: "[T]he process of transmitting, reconciling and, in some cases, confirming payment orders or security transfer instructions prior to settlement, possibly including the netting of instructions and the establishment of final positions for settlement. Sometimes the term is used (imprecisely) to include settlement."

Settlement: "An act that discharges obligations in respect of funds or securities transfers between two or more parties." Also, "the completion of a transaction, wherein the seller transfers securities or financial instruments to the buyer and the buyer transfers money to the seller. A settlement may be final or provisional."

The FPS (U.K.) column in Table 1 separates the settlement (final) stage from other stages because, in most cases, the sender and receiver of funds are not concerned with (and may not even be aware of) when banks settle their own accounts, unless the receiving bank conditions crediting the payee on final settlement between the sending and the receiving banks. As discussed below, the U.K. FPS separates the settlement stage from clearing, so in that system the payer's account is debited and the payee's account is credited within seconds, although banks settle only three times daily.

2.3 "Faster" payments

There is no uniform definition of a "fast payment service."[13] One reason is that the speed of each electronic payment can be measured with respect to at least four steps of the payment process: authorization, clearing, settlement, and notification(s). The first three steps occur in sequence (see rows 1 and 2 of Table 1), whereas notification(s) can be sent to the transacting parties at any stage (or stages) within this sequence.

No funds transfer can be initiated without an authorization, so the first stage is required. However, clearing may be an independent step (as depicted), or it may be combined with (or occur very close in time to) either the authorization stage or the settlement stage. Thus, the following four parameters may be included in the definition of "fast":

1. The ability to process (or at least originate and clear) transactions 24/7/365.
2. The length of time between origination and confirmation of clearing.
3. The length of time between origination and confirmation of settlement.
4. The practice of handling transactions in a nonbatched manner, meaning that each transaction is individually processed through the network (different from the way processing occurs in the existing FedACH in the U.S. and Bacs in the U.K.; see Benson (2009)).[14]

Note that the above four parameters are not mutually exclusive as shown by the fact that the FPS in the U.K. and Singapore seem to satisfy most or all of these criteria.

2.4 Gross versus net settlement

Table 1 separates the settlement (final) stage from all other stages because implementing faster payment services, as done in the U.K., need not rely on instantaneous settlements (which are transfers of funds between two banks via a central bank or a similar clearing house).[15] This implies that the receiving bank may have to extend credit to the payee until settlement is completed. However, a delay in settlement allows banks to aggregate several transactions into a single settlement, and this aggregation may facilitate net settlement, which reduces the amount transferred if banks transact in both directions.

Gross settlements mean one-by-one transfers of funds, which may complicate or overload the network – particularly if the faster payment service results in a high volume of low-value transactions. This suggests one possible explanation of why the FPS process in the U.K. separated the settlement stage from other stages, perhaps to allow banks to gain economies of scale by netting out bi-directional transactions and also to avoid

[13] See FRS (2013) and GPF (2013) for examples.

[14] In a batch payment system, the originating bank bundles several payment requests into a single file that is submitted to the central clearing organization. This explains why faster payment systems may require a technology change.

[15] In the U.S., these bank-to-bank transfers are referred to as wholesale payments.

using CHAPS (the Clearing House Automated Payment System, Britain's real-time gross settlement system).

Instead, the FPS relies on the Bank of England to handle settlement. Thus, the FPS is a new system only for authorization and clearing. Relying on an existing settlement system reduced the construction cost of the FPS and suggests another explanation for why the FPS uses net settlement rather than gross. Note that a delay in settlement creates a tradeoff between the cost of more frequent settlement and the credit risk associated with immediate transfer to the payee.

3. THE U.K. PAYMENT SYSTEM

3.1 The U.K. payment system before the FPS

Prior to the establishment of the FPS in 2008, the payments landscape in the U.K. was similar to that in the U.S. Cash was popular for small transactions, whereas debit cards and credit cards were common for larger-value retail transactions. Checks were also reasonably common and were used for similar purposes as in the U.S. U.K. banks relied on two networks: CHAPS (a real-time gross settlement [RTGS] high-value network similar to Fedwire in the U.S.) and Bacs (formerly known as Bankers' Automated Clearing Services, similar to the automated clearing house (ACH) networks of the Fed and The Clearing House [EPN] in the U.S.), in addition to checks and an ATM network.

Milne and Tang (2005, p. 6) describe Bacs as a provider of three types of payment transactions: bulk (salaries and pension payments, which require submission at least two days in advance of the payment date), direct debit (which are scheduled 14 days in advance), and standing orders (A2A transfers, which require at least two days' notice). Milne and Tang (2005, p. 10) report that immediate person-to-person transfers were most often made using cash or bank drafts. Table 2 roughly compares the payment systems in the U.K. and the U.S. The similarities of the two countries' payment systems suggest that the experience of the U.K. with respect to faster payments may be instructive for the U.S.

3.2 Speed of payment networks in the U.K.

Following is a description of payment methods in the U.K. and an evaluation of the speed at which users can transfer payments.

Cash: if speed is measured by the time it takes for money to change hands, then cash is a fast payment mechanism. If speed is measured as the time it takes to transfer money from one account to another, cash is a slow payment instrument. Two trips to the ATM (or some combination of ATM, bank teller, check cashing store, cash-back at retail, etc.) are required.

Debit: when a consumer initiates a transaction with a debit card, the consumer's bank is immediately informed and typically places a hold on the consumer's account, but the bank does not credit the merchant's account for up to two days [Herbst-Murphy (2013)].[16] This discussion highlights a fundamental difference between card transactions and the operation of the FPS system in the U.K. For card transactions, banks first transfer the money from the issuing bank to the acquiring bank, and only then are funds debited from the sender's account and credited to the receiver's account. In contrast, the U.K. FPS system first debits and credits the payer and payee's accounts, respectively, before the participating banks settle their own accounts with the central bank.

Credit: the credit card market works similarly to the way the debit card market works because Europe uses a dual message system for both credit and debit transactions.[17] However, consumers' billing is

Table 2: Description of payment systems in the U.K. and the U.S.

TYPE	U.K. PAYMENT SYSTEM	U.S. PAYMENT SYSTEM
RTGS (large value)	CHAPS	Fedwire/CHIPS
Batch (slow, any value)	Bacs	FedACH and EPN
Ubiquitous Faster Payment Service	FPS	Not provided by banks
Paper checks	To be phased out	Declining fast
Credit, debit, and prepaid cards	Mostly Chip & PIN	PIN and signature networks and closed loop
Bank account (mainly for bills)	Giro	Bank account number (via ACH)
ATM	Single network	Multiple networks
Coins and notes	British pound	U.S. dollar

Source: Authors' analysis

[16] The U.K.'s Payments Council's Q&A web page states the following: "A debit card transaction will usually be debited from your account on the following working day. However, if the amount of the transaction is above the floor limit of that retailer, the card issuer will earmark the funds on your account at the time the transaction is made. The time it takes for the money to reach the retailer is dependent upon the terms of the contract with their merchant acquirer (bank)." See http://bit.ly/PQEMUu.

[17] Herbst-Murphy (2013) discusses dual and single message systems for debit and credit cards. A dual message system is slower because it was designed for signature credit cards, whereas a single message system relies on a PIN at either a POS or an ATM. Roughly speaking, a single message system combines into a single stage the authorization and the writing of files on the sending and receiving banks. Herbst-Murphy (2013, Figure 1) refers to the stage when accounts are debited and credited as "settlement," whereas in the FPS terminology used in this paper, this stage is referred to as "clearing."

delayed to a predetermined date, or even later if the consumer chooses to borrow by taking advantage of their preauthorized revolving credit. A credit card transaction is revocable. Credit card payments are authorized immediately at the POS and the card issuer is committed to pay at this time.

Paper check: within two days of the day the payee deposits the check into a bank account, the bank must start paying interest on the deposited amount; however, funds may not be available for four days [BIS (2012)].

Bacs: the Bacs system is an electronic system that operates between banks. Consumers do not have direct access to the Bacs system. It is typically used for direct deposit of salary (Bacs direct credits), for paying recurring bills such as utilities (Bacs direct debits), and for business-to-business payments. Before the FPS was implemented, Bacs also was used for payments made via online banking. The Bacs network operates as a batch system. Payments submitted to Bacs are subject to a three-day clearing and processing cycle [BIS (2012)].

CHAPS: CHAPS is a real-time payment system, envisioned for high-value transfers between banks. End-users are charged fees. Transfers executed in CHAPS are irrevocable.

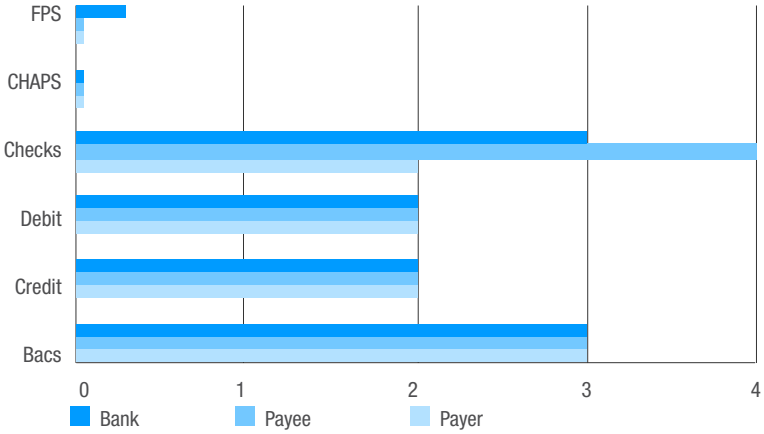
Figure 1 displays rough estimates of the duration of funds transfer from start to end for each payment network from the viewpoints of the sender (payer), receiver (payee), and the participating banks.

As Figure 1 shows, CHAPS and the FPS transfer funds within seconds from the perspectives of both the payer and the payee. However, from the banks' perspective the FPS settles only three times during each business day. Figure 1 also shows the maximum time for payments made via Bacs, payment cards, and checks; however, it is possible that transactions may appear to be faster from the payer's perspective and even from the payee's.

3.3 The U.K. FPS

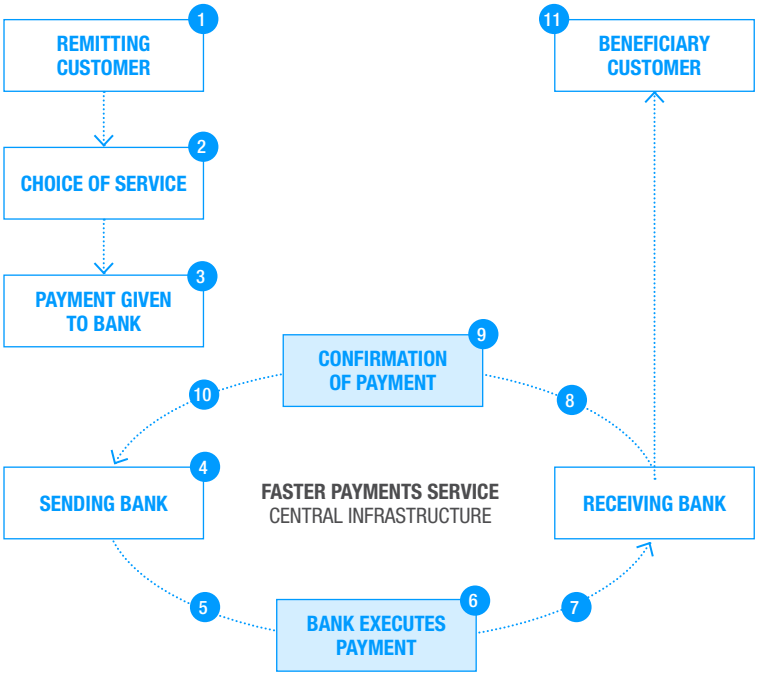
Figure 2 illustrates the structure of the FPS process in the U.K. The sequence of 11 steps illustrated in Figure 2 occurs in a few seconds as follows: (1) A bank customer (payer) decides to send money to a customer of another bank. (2) The payer chooses a mechanism to instruct the bank (mobile phone, online, landline phone, or an ATM). (3) The payer provides the payee's sort (routing) code and bank account number. (4) The sending bank performs security and sufficient funding checks of the payer's account. (5) The sending bank submits the transaction to the FPS. From that stage, the transaction cannot be canceled. (6) The FPS checks that all the relevant information is included and submits the payment instruction to the receiving bank. (7) The receiving bank sends a message back to the FPS that it has accepted or rejected the payment after confirming that the payee's account is valid. (8) The FPS credits the receiving bank (if accepted) and sends a message to

Figure 1: Duration from start to finish from the consumer and bank perspectives



Source: Authors' estimates

Figure 2: How the U.K. FPS works



Source: http://www.fasterpayments.org.uk/about-us/how-faster-payments-works

the sending bank confirming that the transaction was successful (or rejected). (9) The sending bank marks the transaction as complete. (10) The sending bank notifies the payer that the transaction has been completed (or rejected). (11) The receiving bank credits the payee's account for the amount sent.

In terms of speed, the U.K. FPS operates 24/7/365, and clearing and confirmations of individually processed transactions usually occur within a second or two. Settlements are made three times daily; see VocaLink (2009).[18] It is up to the receiving bank to decide to make the funds available immediately to the payee or to delay receipt. In practice, most banks make the funds available immediately. Payments can be originated via the Internet, ATM, over the phone, or via mobile.

Unlike the slower, batch-based networks, the FPS is limited to credit (push) irrevocable transactions.[19] The irrevocable nature of the payment makes correcting errors more difficult than with some other payment methods. While there are mechanisms in place in the U.K. to reverse mistaken or fraudulent transactions, faster payments could be difficult to contest.[20] If a payer provides the wrong sort code or account number when making a payment, the bank must make a reasonable effort to recover the money, but the bank is not liable for losses.

By late 2014, 49 million account holders in the U.K. (compared to an adult population of 52 million) had access to the FPS. Initially, each transfer was limited to £10,000 ($15,365). Some banks have raised the limit beyond £10,000 for individual customers to £250,000 for business customers.[21] FPS values accounted for 1.0% of total clearing values in 2013 [PC (2014a)]. This share by value is low due in part to the size of other types of transactions. For example, the average CHAPS transaction was £1.99 million in 2014, while the average FPS transaction was £589, as discussed below and shown in Figure 9.

Through 2014, the introduction of the FPS in the U.K. has had little or no effect on transactions made at the POS. For purposes of this discussion, POS refers to payments that must be made prior to the delivery of goods. An FPS system announced for 2015 enables users to pay directly from their bank accounts by scanning a barcode or tapping an NFC reader with their mobile phones.[22]

One improvement to the FPS was the introduction of mobile FPS, whereby users who register their accounts can make payments using their mobile phone numbers without having to reveal their bank account details.[23] This service aims to make it easier for individuals to pay one another.

[18] Because settlements occur only three times per day, banks in the U.K. have signed a loss-sharing agreement in case one of the banks fails before funds are settled.
[19] The newly constructed faster payments system in Singapore (called Fast, for Fast And Secure Transfers) is able to handle debit requests: http://bit.ly/2GYFw4H.
[20] "Bank digit mistakes costly" BBC One (June 19, 2013), http://bbc.in/2F7D0wm
[21] See http://bit.ly/2l0uEVn
[22] Rossi, B., 2013, "VocaLink announces new mobile payment system 'Zapp,'" Information Age, June 25, http://bit.ly/2teUstt
[23] The service is called Paym; see http://bit.ly/1igZzbS. Like FPS, Paym is offered by the participating banks, which guarantee that 90% of bank customers will have immediate access to this service.

A second enhancement is adherence to international standards, which eventually would permit faster payments between countries. The U.K. FPS and Singapore's Fast are compatible with ISO 20022.[24] The purpose of this standard is to unify payment messages across all electronic payment systems in Europe and all other participating countries.[25] Concerns have been raised that payment systems in the U.S. are incompatible with ISO 20022.[26]

4. USES AND POTENTIAL BENEFITS OF FASTER PAYMENTS IN THE U.K.

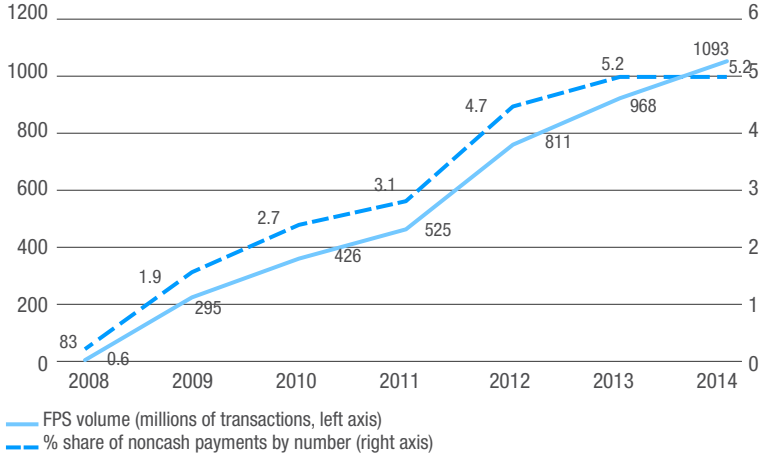
Using limited data available, this section describes how businesses and consumers in the U.K. used the FPS shortly after its introduction. Data are limited, but there are a few interesting findings. In 2014, 1.1 billion payments initiated by consumers, businesses, or government were processed via FPS.

Overall in the U.K. in 2013, consumers, businesses, and government made 18.5 billion noncash payments, so payments via the FPS represented about 5.2% of all noncash payments by number (up from 4.7% in 2012). Including cash payments, faster payments were about 2.6% of all payments by number (PC 2014b, Table 27.1). Figure 3 displays the volume of FPS transactions since 2008, when the FPS became operative.

As shown in Figure 4, faster payments in the U.K. consist primarily of three types of payments:

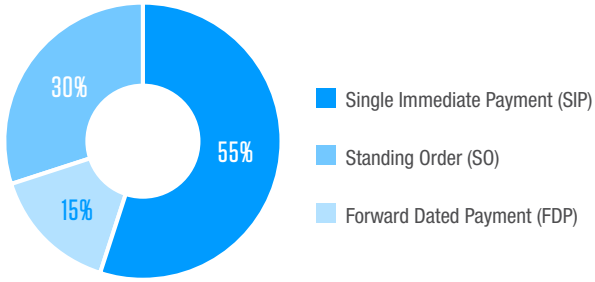
- **Single immediate payment (SIP):** A one-time payment initiated via Internet banking, telephone banking, or an ATM, to be executed immediately. For example, a consumer might use an SIP to pay a credit card bill. In May 2014, SIPs were the dominant type of faster payment by volume: 55% of all FPS transactions by number.
- **Forward-dated payment (FDP):** An instruction to a bank to make a one-time payment on a future date. For example, a business or consumer might schedule a tax payment due on a future date. In May 2014, FDPs were 15% of all FPS transactions by number.
- **Standing order (SO):** Regular recurring payments for a set amount, to be made on the same day of every month or week. For example, a business might schedule a monthly payment to a cleaning service. Standing orders can be set up at any time, but this payment type is only sent Monday through Friday. In May 2014, SOs represented 30% of all FPS transactions by number [FP (2013), updated using "Payments Statistics Monthly"].

Figure 3: FPS volume, 2008–2014 (in millions)



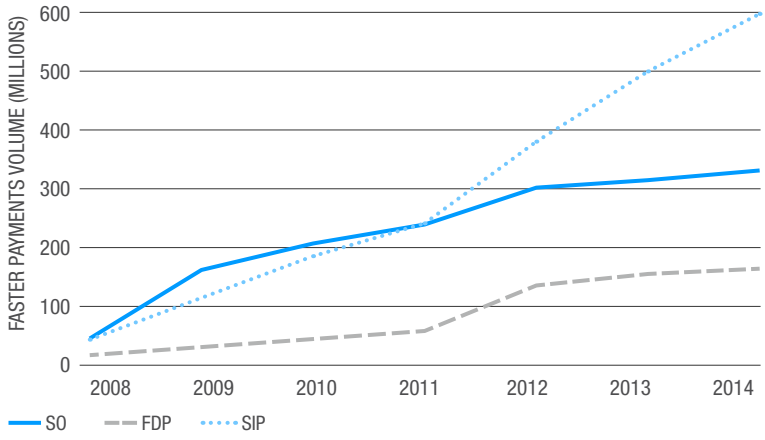
Source: PC 2014b, p. 14, Table 1.6, and authors' calculations based on PC 2014b, p. 85, Table 27.1

Figure 4: Shares of faster payments types (by number of payments) in the U.K., May 2014



Source: FP (2013), updated using "Payments Statistics Monthly"

Figure 5: FPS volume, 2008–2014 (in millions)



Source: PC (2014b, p. 14, Table 1.6)

[24] See, http://bit.ly/2oFzM9g.
[25] See, European Payment Council, http://bit.ly/2CYjxIK.
[26] See, http://bit.ly/2tbIMIW.

Previously, these types of payments initiated via telephone or online banking were executed via the Bacs system, where the payer initiated the payment on business day 1 and the payee received the payment on business day 3. At best, the money would reach the beneficiary two days later (for a payment made on Monday the beneficiary would be credited on Wednesday), but since this was a Monday-through-Friday service and had an evening cut-off time, a payment initiated on a Friday evening would reach the beneficiary on the following Wednesday. For all FPS, including the almost half of these faster payments scheduled in advance (Figure 4), this multi-day timeline from initiation to receipt no longer applies.

Figure 5 shows the strong growth between 2011 and 2014 in SIP, that is, one-time payments authorized online or by phone or ATM. The number of SIP and FDP made using the FPS grew by more than 75% in 2012 ([PC (2013c), p. 82].

As noted above, consumers, businesses, and governments use the FPS. Much of the analysis in this section relies on the "2013 Faster Payments Service Traffic Survey" [FP (2013)], which provides data on all transactions over the FPS for five dates in May 2013.[27] Unfortunately, the report (discontinued in later years) does not provide full information about the payer or payee of the transactions and provides only very limited information about the transaction. It makes use of sort codes, which are codes associated with each transaction that primarily identify the bank. As it happens, knowledgeable sources are able to recognize some sort codes as being associated with banks that particularly specialize in businesses, consumers, or governments, and this provides some information. But sort codes cannot be used to identify the specific payer or payee.

4.1 Consumer payments

According to the FP (2013, p. 23), consumers made an estimated 487 million payments, or approximately 60% of the payments sent via the FPS in 2012. Another data source, a survey of U.K. consumers, found a smaller number of faster payments by consumers in 2012, 356 million [PC (2013b)]. Figure 6 shows how the U.K. consumer payment survey classifies consumer payments.

4.1.1 ONE-TIME PAYMENTS BY CONSUMERS

In the U.K. in 2012, consumers made almost 31 billion payments, including 25.9 billion "spontaneous" payments, which the Payments Council defines to include purchases in person, by mail, and online; payments for services, for example at hotels and restaurants; and payments to individuals [PC (2013b)].[28] Spontaneous payments also include one-time credit card payments. Of the 25.9 billion spontaneous payments, 239 million (1%) were electronic payments.

The most common type of payment made via the FPS is payment of a credit card bill [FP (2013), p. 23–24]. In 2012, more than two-thirds of all credit card bill payments were made using FPS [FP (2013), p. 35]. This suggests that consumers are taking advantage of same-day receipt to pay credit card bills on time with the most up-to-date knowledge of their financial situation, which they could not have done previously with SIP.

By value, the largest total value amount of FDPs according to the Faster Payments Service Traffic Survey (2013) was paid to public sector sort codes. FDPs also are one-time payments. This suggests that consumers are using FDPs to schedule the payment of taxes.

4.1.2 REGULAR PAYMENTS BY CONSUMERS

In 2012, U.K. consumers also made 4.8 billion payments for "regular," or recurring, commitments, including household expenses like rent, gasoline, and insurance, and personal commitments like health insurance, subscriptions, and loan repayments. Of these recurring consumer payments in 2012, 3.6 billion (75%) were electronic payments [PC (2013b)].

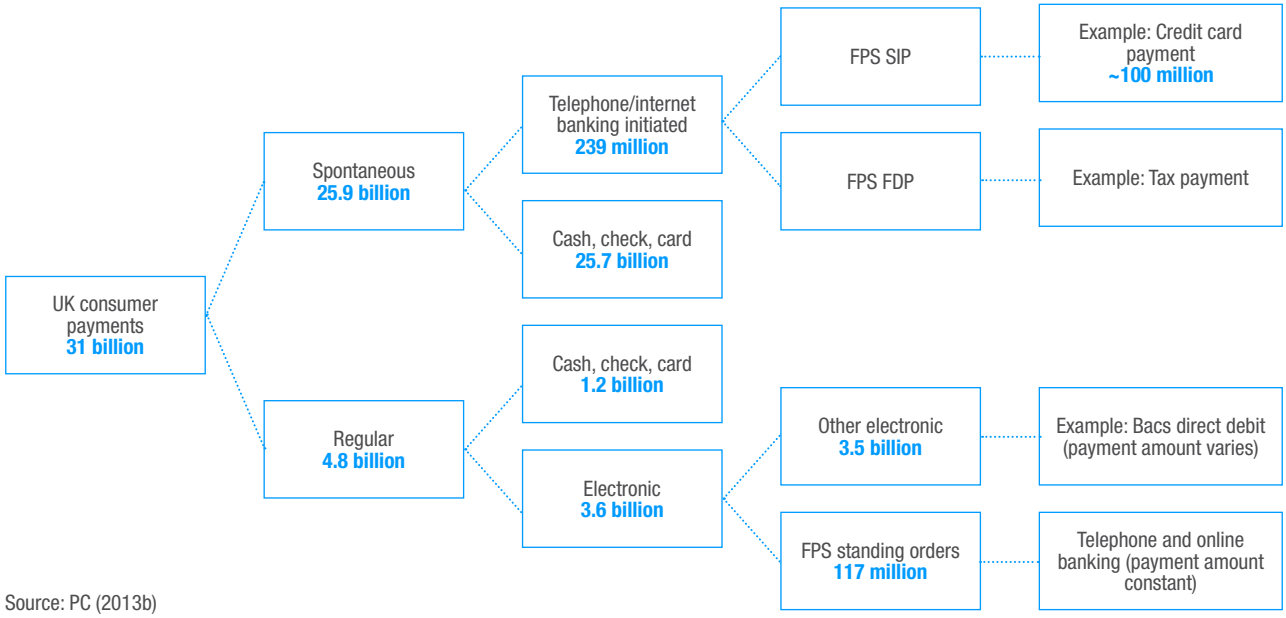
Almost all constant-value recurring payments authorized by telephone or online are executed via FPS standing orders (nonparticipating banks would be the exception): 94.6% [PC (2013a)]. The remainder are processed via Bacs. Thus, the FPS system has almost entirely taken over the standing order market.

PC (2013b) reports 117 million payments by FPS

[27] This annual survey reports on all transaction data for five dates in May 2013, month-end and month-start (April 30–May 1), a mid-month weekend [Friday, May 17 (encompassing May 18 and 19 because these payments settled on Monday May 20)], and two days around the middle of the month (May 21–22).

[28] The Faster Payments Tracking Survey was discontinued in later years. This section uses 2012 data for comparability.

Figure 6: Consumer reported payments in the U.K. in 2012



Source: PC (2013b)

standing orders in 2012. FP (2013) arrived at a similar estimate for recurring payments made by telephone or online. Use statistics for the FPS in May 2013 show that about one-third of all payments associated with individuals were made on the last day of the month. Of payments by individuals on the last day of April and first day of May, 80% were standing orders [FP (2013, p. 25)]. That is, about one-quarter of the use of FPS by individuals (129 million payments) is for SOs at the end of the month, most likely for recurring monthly bills for constant amounts.

According to FP (2013), FPS SO payments are for lower values than FPS one-time payments. More than half of SOs at both the beginning of the month and the middle of the month were for £100 or less. On all days surveyed in May 2013, about 20% of SOs were for £10 or less, "many for £4.33 and £4.34 exactly, probably monthly payments of £1 weekly commitments" [FP (2013), p. 21]. This suggests that the FPS is being used to automate small payments by individuals to other individuals, for example, for workplace coffee clubs or lottery pools.

An important consideration is that if the payment level varies from month to month, such as with a typical telephone bill, automatic payment cannot be done by SO. Consequently, constant-value payments, for example, for rent or life insurance, are much more likely to be processed over the FPS than are payments for, say, electric utility bills, which vary in value from month to month. This is an important limitation of the FPS as implemented in the U.K.[29]

4.2 Business and government

In the U.K. in 2012, businesses made 3.5 billion payments [PC (2013c), p. 87]. Approximately 324 million business (including government) payments (calculated as 40% of FPS by number), or 9% of business payments by number, were made via the FPS.

4.2.1 BUSINESS-TO-BUSINESS

For business-to-business payments, PC (2013d) reports that of the 832 million automated payments in 2012 (made by Bacs direct debit or direct credit, standing order, FDP, and SIP), 60 million (about 7%) were FDPs or SIPs via the FPS [PC (2013d), pp. 39-40]. Small and medium-sized businesses may now choose to receive credit and debit card merchant payments via the FPS (reducing settlement time by as much as three days) [PC (2013d), p. 46].

[29] In practice, direct debits make up the majority of recurring, scheduled payments by consumers.

4.2.2 BUSINESS-TO-CONSUMER

Business-to-consumer payments were about 264 million by number in 2012. For business-to-consumer payments, FP (2013) found that sort codes associated with businesses sent more payments on Friday, May 17, than on any of the other four days of the survey. "This was in part driven by employment agencies and payroll companies making weekly wage payments" [FP (2013), p. 15]. This suggests a potential benefit of faster payments. When the payment of wages and salaries can be based on contemporaneous data on employment status or hours worked, payments are more accurate and timely. FP (2013, p. 14) reports "Employment agencies paying staff on weekly bases using the Faster Payments Service tend to use single immediate payments because of its flexibility." In addition, government can use the FPS to pay benefits to recipients. As of 2013, however, most workers are paid wages and salary via Bacs Direct Credit. In 2013, 90% of working adults in the U.K. were paid by Bacs Direct Credit [PC (2013), p. 31].

According to the FPS, financial users and businesses users were the second and third largest users, respectively, of SIPs. Some of these financial users and businesses seem to be using speed of payment as a competitive advantage. FP (2013, p. 27) commented on the growing use of faster payments by businesses to pay customer refunds and insurance claims. The Payments Council [PC (2013), p. 34] notes that "a number of new businesses have emerged in recent years advertising speedy payment," citing businesses that buy second-hand cars or jewelry, lenders, and gaming companies.

Financial firms also use SIPs. "Most payments by financial firms to individuals were around £100 or less and included a relatively large number of payments of less than £1. As noted earlier in this report, these could be interest payments from old savings accounts" [FP (2013, p. 27)].

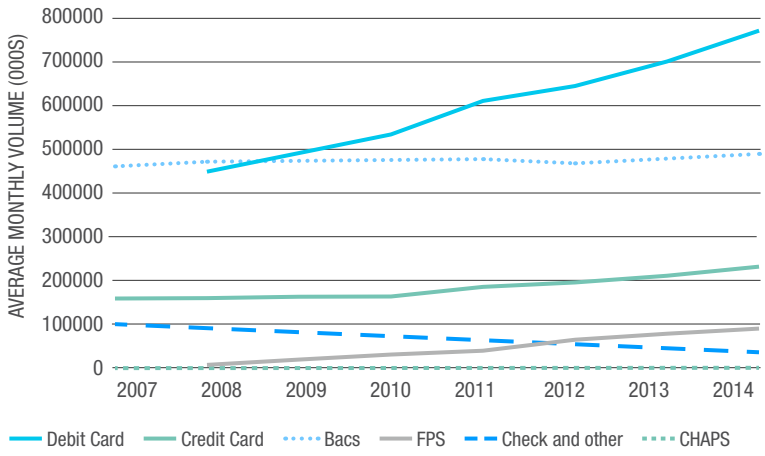
4.3 Consumers and the U.K. FPS

Research by the U.K. FPS found that person-to-person payments for coffee, lunch, shared housing expenses, etc., amount to £12.6 billion per year [PC (2014c)]. At the time of this research, it was anticipated that the Paym mobile method, introduced in mid-2013, would make it easier for friends and family to settle IOUs.[30]

As of 2016, the U.K. FPS offered technology for POS transactions.[31] Research has shown that speed at checkout is a relevant consideration [Klee (2008); Schuh and Stavins (2015)]; therefore, a payment method that slows down speed at checkout is not attractive. As Figure 7 illustrates, debit card volume increased after the introduction of FPS. At the time of writing this article, costs to merchants for POS services are unknown. If costs prove smaller than merchant card fees (including interchange), merchants could surcharge consumers for the difference between the costs of accepting the two payment methods, as permitted by U.K. consumer protection rules. Discounts from the stated price for the use of a particular means of payment are permitted [BIS (2013)].

Broader adoption of international standards such as the ISO 20022, the standard for financial services messaging, could facilitate the use of faster payments for cross-border payments. VocaLink suggested some benefits of international standards for cross-border payments. "[T]he standardisation of approach reduces the burden of interoperability between systems, assisting both reconciliation and integration with the end to end business process, as well as enabling a greater "payload" of identifying information to accompany the payment" [VocaLink (2013, p. 9)]. ISO 20022 includes standards for payment initiation, cancellation, and modification of payments, and settlement instructions.[32]

Figure 7: U.K. transaction volumes by payment method, 2007–2014 (before and after the FPS)



Source: PC (2013c), updated

[30] Faster Payments, "Paym launch confirmed – pay using just a mobile number from 29th April" http://bit.ly/2I3WVtW.

[31] FinTech Futures, 2016, "VocaLink, Zapp bring faster payments to retailers," BankingTech, June 13, http://bit.ly/2FpEJg9 and Rossi, B., 2013, "VocaLink announces new mobile payment system 'Zapp'," Information Age, June 25, http://bit.ly/2FReZXy.

[32] ISO 20022 Payments Dashboard Business Processes Description, http://bit.ly/2Fb5k10

Overall, from a consumer perspective, it appears that consumers in retail settings in the U.K. and the U.S. have good options for fast payments. Debit and credit transactions appear immediately, and cash is often an option. Merchants may see this differently, as their payment may be delayed. But consumers tend to be the driving decision-makers in retail settings, and they have little reason to adopt something new, unless incentives change; for example, if merchants were to choose to offer discounts. However, person-to-person transactions are different. These are often completed by check, a slow process that often involves physically mailing a check or depositing at a bank, ATM, or via the Internet (by taking an image of the check). Similarly, real-time payments may be attractive in bill-pay contexts. Unlike the case with ACH or check payments, a consumer can schedule a real-time payment at the last minute, which supports better money management (and procrastination).

5. COSTS OF THE FPS IN THE U.K.

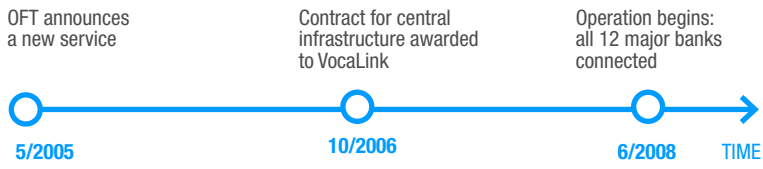
The U.K. FPS experience provides a good example of how a general-purpose, fast payment system can be constructed and become operational in three years. Figure 8 below illustrates the construction timeline.

The FPS started operating in 2008. The key to its success is that commercial banks had a strong incentive to construct and connect to such a network. The whole process was pushed forward by the Office of Fair Trading (OFT, one of U.K.'s antitrust authorities), which offered commercial banks no choice but to remove the float from funds transfers [VocaLink (2009)]. In addition, at that time, check clearing in the U.K. was planned to be phased out in October 2018 (since then postponed).[33]

5.1 The U.K. payment system investment decision

To determine whether the benefits from an enhanced payment system outweigh the costs requires the decision-makers to examine various technological issues in general and all the available options related to existing electronic payment networks in particular. This is because the choice of technology has a direct impact on both the expected benefits and the expected costs. In general, such a debate would focus on four options:

Figure 8: Timeline for the construction of the U.K. FPS.



Source: Faster Payments Service

1. Speeding up an existing A2A payment system; for example, making the Bacs (ACH batch based) clear transactions several times during a 24-hour cycle or modifying CHAPS to permit low-value transactions.

2. Building a totally new A2A faster payment system, such as the VocaLink FPS.

3. Modifying other existing payment networks to make them suitable for A2A transfers; for example, using an existing debit card or an ATM network.

4. Using an existing Real Time Gross Settlement (RTGS) network by reducing end-user fees for low-value funds transfers.

The costs of establishing and maintaining a faster payment service involve three main components:

1. The installation cost of constructing, deploying, and maintaining the central infrastructure of the FPS.

2. The connection cost to each individual bank of adopting new technology and capital to access the new fast payment network.

3. The transfer costs of possible reductions in the revenue of banks and nonbank money transmitters resulting from shifting some volume from other payment services to the new FPS. (Lost revenue is not a social cost; rather, it is a transfer from one agent in an economy to another; see section 5.4.)

The cost estimates here are from sources at VocaLink, which runs the central infrastructure and also conducts surveys of participating banks in order to learn about the cost to banks. At the time of this writing, VocaLink (subsequently acquired by MasterCard) was 100% owned by banks. Sources at VocaLink have indicated that the cost to build and launch the U.K. FPS, plus the operation for the life of the initial contract (seven years), is estimated at somewhere between £150 and £200 million (U.S.$230–U.S.$307 million), of

[33] http://bit.ly/2Ff5xA4

which £40–£50 million (U.S.$61–U.S.$77 million) was a fixed cost paid up front by the 12 participating banks for the construction and launch. These figures do not take into account individual bank costs. The costs to individual banks were wide ranging, depending upon the bank's existing capabilities and the extent of the changes each bank elected to include within the scope of its FPS project.

According to VocaLink, the cost of constructing Singapore's Fast was lower, due to experience with the U.K. system. It should be mentioned that VocaLink does not bear any volume risk because it does not charge banks any per transaction fee. That is, banks pay a flat fee to use the service, so the system is immune to demand fluctuations. Because it operates below capacity, volume also does not affect cost.

As for the second cost component, the adoption cost to each participating bank varied significantly among the banks, even when adjusted for volume. Some banks used this opportunity to overhaul their entire accounting system in order to accommodate fast clearing. Some banks reported this cost to be in the hundreds of thousands of British pounds (rather than in the millions; see VocaLink (2009), p. 16]. On the other hand, some banks spent in the tens of millions of British pounds. The major problem in estimating bank-specific adoption cost is that for banks that overhaul their entire accounting system, it is difficult to isolate the portion that is attributable solely to the adoption of the FPS. Since banks are not charged any per transaction fee, any increase in volume does not add to a bank's total cost.

Table 3 provides a summary of the "real" overall cost of the implementation in the U.K. The term "real" refers here to the diversion of human and physical resources from other activities. These can also be viewed as "social costs."[34]

We have already noted that the costs of FPS are very small relative to U.K. GDP (national income). Another way to evaluate the relative total cost of FPS is compare it to the per capita value of benefits required. With a U.K. population of 63 million people and estimated maximum total cost of £800 million (U.S$1.23 billion), the FPS would require a per capita annual benefit of £2.05 (U.S.$3.15) to give the seven-year investment project a positive net present value.[35] For example, if the FPS helped avoid a late fee on one monthly bill per consumer per year, it would more than amortize the total costs.

Table 3: Estimated cost of building and maintaining FPS in the U.K.

BEARER	COST DESCRIPTION (REAL)	ESTIMATED AMOUNT, MIN TO MAX
Split by 12 banks	Central infrastructure: construction (fixed cost)	£40 million–£50 million (U.S.$61 million–U.S.$77 million)
Split by 12 banks	Central infrastructure: Maintenance (variable cost)	£100 million–£150 million (U.S.$154 million–U.S.$230 million), spread over seven years between 2008 and 2015
Each of 12 banks	Adoption costs	£0.10 million–£50 million (U.S.$0.15 million–U.S.$77 million); max times 12 banks = £600 million (U.S.$922 million)

Source: VocaLink representatives in email and phone conversations with authors.

5.3 Changes in float

In assessing the impact of the FPS process in the U.K. on the costs or gains from float, Milne and Tang (2005, p. 7) point out that a transition from Bacs (ACH) to FPS will not have any float-related impact on the sender and the receiver because even under the slow system, "the debiting and crediting of customer bank accounts takes place on the same day so there is no float income for banks generated by either of these payment instruments." This statement refers to "bulk credits," which are transfers such as salaries and pensions, as well as "bulk debits," which are payments for utility or other variable-amount recurring bills (as noted above, these variable-amount, recurring payments are not processed by the FPS). However, float can arise with standing orders (such as regular payments for magazine subscriptions and club dues), "where it is usual for banks to debit customer accounts two working days before the crediting of the recipient account. However, recently some individual banks have changed their practice… so this change in practice eliminates float on standing orders paid by customers of the bank."

5.4 Revenue impact on existing payment networks

The question of whether the new service would generate substitution from other payment instruments, such as checks, cash, CHAPS, Bacs, and cards, was raised in a preliminary study commissioned by the OFT [Milne and Tang (2005), p. 16]. That study mentioned that a large portion of scheduled payment orders should be unaffected by the new service because they are scheduled in advance for fulfillment at a certain future date; such payments include salaries, utility bills, and pension payments.

[34] Gains or losses from float are not included in this table, as they are generally netted out in general equilibrium.
[35] Calculated using a discount rate of 3%.

Figure 7, above [computed from PC (2013c)], confirms that the increase in the volume share of the FPS transactions did not correspond to declines in the volume of CHAPS, Bacs, and debit card transactions. The only significant reduction in payments was in the volume of checks, which were scheduled to be phased out in the U.K. Because checks are also used for person-to-person transfers, the FPS may have affected check volume.

Figure 9 displays average transaction values from the time that the FPS was introduced in the U.K. through 2014. Note that the CHAPS average transaction value in 2014 was £1.99 million, which would fall above the vertical axis limit.

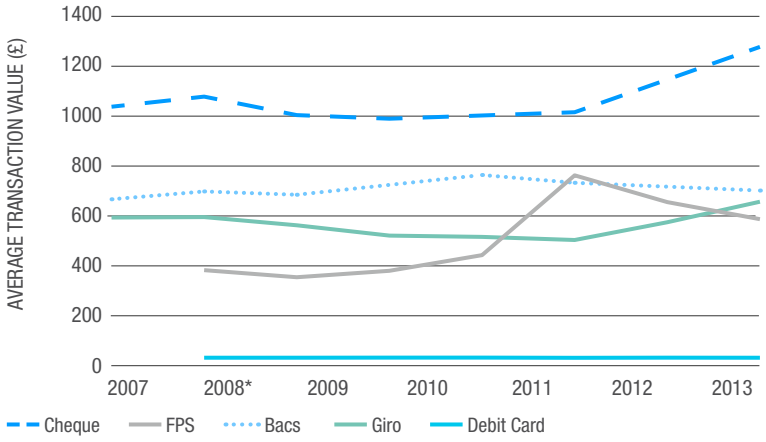
The figure reveals a sharp increase in transaction values made via the FPS in 2012, partly because participating banks increased the limit on the amount that could be sent. Note that none of the other payment methods exhibited major change.

The volumes of Bacs and CHAPS have not decreased appreciably (Figure 7), but it is not possible to say how the introduction of FPS has affected these two electronic networks. A comprehensive model of the payment system is needed to properly estimate substitution among payment methods. For the sake of illustration, one way to approach this kind of computation would be to look at a potential loss of revenue to the banks if some volume from CHAPS switched to the FPS, which currently does not charge payers and payees.[36] The CHAPS volume in 2012 was 33,936,000. Banks charge £30 (U.S.$43) to send (they do not seem to charge for receiving). The median CHAPS transaction value in 2003 was £17,000 (U.S.$27,086).[37] If a £10,000 (U.S.$15,365) restriction corresponded to 25% of the distribution, then the maximum revenue loss would be 1.018/4 = £0.255 billion (U.S.$0.362 billion) Based on this rough calculation, a potential loss to banks could be in the range of £0 to £0.255 billion (U.S.$0 to U.S.$0.362 billion).

5.5 Future costs

Table 3 shows the estimated maintenance cost of the FPS in the U.K. to be in the range of £100–£150 million (U.S.$154–U.S.$230 million) for the entire first seven years of operation (2008–2015). This cost is likely to continue in the future because, so far, the capacity of the network seems sufficient for current traffic. This cost was divided among the 12 banks that started and owned the project since the beginning. The cost per bank falls as more banks join.

Figure 9: U.K. transaction values by payment method, 2007–2014 (before and after the FPS)



*Note that for FPS, 2008 only includes July – December
Source: PC (2013c), updated

However, future enhancements that will use the FPS, such as POS applications, may incur additional costs. For example, the Paym mobile service also contains a user directory so that the consumer whose account is credited does not have to reveal his bank account to the sender. It is natural to assume that some add-on service of this type could also be provided by nonbanks, such as merchant organizations and merchants who adopt the FPS, who then could bear some cost of subsequent enhancements.

5.6 Revenues

Since operation began, banks in the U.K. have provided FPS to their customers free of charge. Consequently, no revenue has been collected. There may be two reasons for this. First, as with most new networks, to gain momentum, charging no fees could be viewed as providing "introductory offers" to end users so they could assess the gains from using the FPS. Second, the participating banks themselves were not charged any per transaction fees to use the FPS, only fixed fees that were spread over the first seven years of operation. Banks have, therefore, borne zero marginal cost (the cost of making one additional FPS transaction). This means that if banks were to charge end-users per transaction fees, the basis of the fee could not be marginal cost, but rather "demand" or "utility."

[36] Due to lack of data on revenue generated by U.K. banks from Bacs services, similar estimates of potential loss of revenue from shifting volume from Bacs to FPS are unavailable.
[37] See Bank of England Archive, http://bit.ly/2FezZKY.

In the future, banks will have to decide whether to charge users nominal fees that would cover their initial investment and operating cost or whether to cross subsidize this service. VocaLink (2009, p. 4) has already questioned banks on their vision for future revenue, and the response received has been as follows: "Two-thirds of banks interviewed were very positive that Faster Payments could deliver new revenue streams, with potential revenues identified in the business-to-consumer segment reaching £2.9 billion by 2018 and £1.9 billion in the business-to-business space."

6. CONCLUSIONS AND IMPLICATIONS

Faced with a directive from the U.K. Office of Fair Trading to increase the speed of payments, the U.K. banking system chose to build a totally new A2A faster payment system. By most counts, the resulting VocaLink FPS has been a success. In addition to being faster, it is low cost and technologically advanced relative to legacy payment systems, features that benefit U.K. consumers and firms. In particular, the U.K. FPS provides a valuable service to customers with a pressing need to make exceptionally fast payments (relative to legacy systems) in certain payment contexts that help them avoid fees or other costs, pecuniary and nonpecuniary.

Given this success, it is somewhat surprising that in 2015, seven years after starting to operate in the U.K., FPS accounted for a low share of total U.K. payments (around five percent).[38] Whether this outcome was the result of intentional planning and design of the FPS, deficient demand for the service, resistance to revenue losses, or some other reason is not known at this time. However, it appears at least somewhat paradoxical that a lower cost, technologically advanced, faster payment system would not have spread more quickly and widely. Studying the subsequent performance of the U.K. FPS will be an important topic for future research and for understanding the ultimate value of such systems for modern economies.

Because the U.K. VocaLink system was an early successful application of faster payments, it has implications for subsequent efforts to speed up payment systems in other countries. For example, VocaLink reported to the authors that its U.K. experience yielded many insights that aided its development of a similar system in Singapore. In 2012, the Federal Reserve announced a desire to improve the speed, security, and end-user experiences in the U.S. payment system within a decade. Since then, the Fed has encouraged the U.S. payments industry to develop, propose, and install a faster payment system(s). Given the marked similarities between the U.K. and U.S. payment systems, the potential benefits of faster U.S. payments are likely to be similar to those experienced in the U.K.

However, there are differences between the U.S. and U.K. economies that may affect the costs and design of a U.S. FPS. Chief among these differences may be the structure of the banking systems. There are many more banks in the U.S. In 2014, the top five U.K. banks held 98% of all deposits, whereas the top five U.S. banks held only 56%. Another potentially important difference is how revenues would be raised to pay for the FSP services to be provided. In the U.K. FPS, no fees were charged directly to users, unlike payment card schemes that typically impose fees on payees (usually merchants) but not payers (typically consumers). Thus, the optimal business strategy for long-run success in a U.S. FPS may differ from that of the U.K. FPS.

In any case, the U.S. payments industry and policymakers have a distinct advantage in their decision-making because they can study and learn from the experiences of the U.K. and other countries with faster payment services. Studying the evolution of faster payment systems in various countries over time will also be an important line of future research.

[38] 2016 UK Payment Markets – Summary available at http://bit.ly/27SzIli

References

Benson, C., 2009, "A look at the UK's faster payments service," Paymentsviews

BIS, 2003, "A glossary of terms used in payments and settlement systems," Bank for International Settlements

BIS, 2012, "Statistics on payment, clearing and settlement systems in the U.K.," Bank for International Settlements

BIS, 2013, "BIS guidance on the consumer protection (payment surcharges) regulations 2012," Department for Business Innovation & Skills, Bank for International Settlements

Biehl, A., J. McAndrews, and C. Stefanadis, 2002, "A review of the retail and wholesale markets for funds transfers," working paper, Payments Studies Function, Federal Reserve Bank of New York

Borzekowski, R., and E. K. Kiser, 2008, "The choice at the checkout: quantifying demand across payment instruments," International Journal of Industrial Organization 26:4, 889–902

Clear2Pay, 2014, "Flavors of fast," http://bit.ly/2CZVsRy

FRS, 2002, "The future of retail electronic payments systems: industry interview and analysis," Staff Study 175, Board of Governors of the Federal Reserve System

FRS, 2013, "Payment system improvement - public consultation paper," Board of Governors of the Federal Reserve System

FRS, 2014, "Payment system improvement – public consultation paper industry feedback summary," Board of Governors of the Federal Reserve System

FP, 2013, "2013 Faster Payments Service Traffic Survey," Faster Payments

Herbst-Murphy, S., 2013, "Clearing and settlement of interbank card transactions: a MasterCard tutorial for Federal Reserve payments analysts," Payment Cards Center, Federal Reserve Bank of Philadelphia Discussion Paper

Jacob, K., and K. Wells, 2011, "Evaluating the potential of immediate funds transfer for general-purpose payments in the United States," Chicago Fed Letter No. 292a.

Klee, E., 2008, "How people pay: evidence from grocery store data," Journal of Monetary Economics 55:3, 526–541

Koulayev, S., M. Rysman, S. Schuh, and J. Stavins, 2012, "Explaining adoption and use of payment instruments by U.S. consumers." Federal Reserve Bank of Boston working paper no. 12-14

Laffont, J-J., and J. Tirole, 2002, Competition in telecommunication, The MIT Press

Lodge, G., 2014, "Real-time payments: dispelling the myths," Celent, http://bit.ly/2oS3qHt

Milne, A., and L. Tang, 2005, "An economic analysis of the potential benefits and dis-benefits of faster payments clearing," Office of Fair Trading No. 795

GPF, 2013, "What will the role of bank accounts be as payments evolve?" NACHA's Global Payments Forum, http://bit.ly/2D0R1WM

Payments Statistics Monthly, available at fasterpayments.org.uk

PC, 2013a, "Free industry statistics," Payments Council

PC, 2013b, "UK consumer payments," Payments Council

PC, 2013c, "UK payment statistics," Payments Council

PC, 2013d, "UK payment markets," Payments Council

PC, 2014a, "Free industry statistics," Payments Council

PC, 2014b, "UK payment statistics," Payments Council

PC, 2014c, "IOU UK," Payments Council

Polasik, M., J. Górka, G. Wilczewski, J. Kunkowski, K. Przenajkowska, and N. Tetkowska, 2013, "Time efficiency of point-of-sale payment methods: empirical results for cash, cards and mobile payments," Enterprise Information Systems 141, 306–320

RBA, 2012, "Strategic review of innovation in the payments system: conclusions," Reserve Bank of Australia

Schuh, S., and J. Stavins, 2015, "How does security and speed influence consumers' payment choices?" Federal Reserve Bank of Boston Working Paper 2015-1

Shy, O., 2012, "Account-to-account electronic money transfers: recent developments in the United States." Review of Network Economics 11:1, 1–23

Stavins, J., 1997, "A comparison of social costs and benefits of paper check presentment and ECP with truncation," New England Economic Review, Federal Reserve Bank of Boston, July/August

Summers, B., 2012, "Payment systems: design, governance and oversight," Central Banking Publications

Summers, B., and K. Wells, 2011, "Emergence of immediate funds transfer as a general-purpose means of payment," Federal Reserve Bank of Chicago, Economic Perspectives, vol. 35, 3rd Quarter

TCH, 2013, "US payment system: recommendations for safe evolution and future improvements," The Clearing House

Tirole, J., 1989, The theory of industrial organization, MIT Press

VocaLink, 2009, "Tomorrow happened yesterday," VocaLink and PriceWaterhouseCoopers publication

VocaLink, 2013, "Federal Reserve Payment System improvement public consultation paper VocaLink response," http://bit.ly/2tfofCp

Household deformation trumps demand management policy in the 21st century

IORDANIS KARAGIANNIDIS | Associate Professor of Finance,
The Tommy and Victoria Baker School of Business, The Citadel

D. SYKES WILFORD | Hipp Chair Professor of Business and Finance,
The Tommy and Victoria Baker School of Business, The Citadel

ABSTRACT

Demographic impacts that will disrupt traditional demand management policy tools are examined. Given the demographics of aging, the lifecycle of consumption for a country, as well as an individual, this paper concludes that one of the key drivers of demand management policy will disappear from many of the wealthy economies over the next 30 years. Economists often speak of the liquidity trap justifying fiscal stimulus. The new mantra may become "forget the liquidity trap, it's the demographic trap" that weighs down the economy. As a result, systemic issues will loom large, affecting housing demand, development models, and portfolio valuations for many of the pension funds needed to support an aging population. Narrowly focused upon household (de)formation, the paper's analysis allows to draw wider implications of the impact of the aging populations.

1. INTRODUCTION

President William Jefferson Clinton was famous for the term "It's the economy, stupid," and Lord Keynes famously coined the term "the liquidity trap." With both in mind, politicians can justify spending to ensure sufficient demand. Even without reference to either term, central bankers can justify zero interest rates to stimulate demand, for the government to save us from slow growth, and thus stimulate us to buy more big screen televisions (only partially sarcastic). To those who believe we are all dead in the long run or that monetary policy can stimulate real growth via interest rate demand management policy, this paper will accept these concepts as truth, but only as a starting point. An essential premise of these types of government interventions to stimulate demand is that people will spend the money (much like Nancy Pelosi's comment concerning the benefit of greater welfare spending to help increase employment), thereby stimulating demand via the multiplier that will eventually cause greater real growth as the consumer consumes more. This leads to greater production, more jobs, and brings the economy out of the liquidity trap. Shop till we drop will solve all evils.

There is one glaring problem. The demographics of the 21st century make this type of policy less predictable, more likely to fail, and could lead to distortions much greater than expected by the central planners manipulating the system. Older populations do not buy as much "new stuff." They are selling old stuff, downsizing, using auction sites and dispensing with stuff, and thus not net accumulating. As such, stimulating demand via a policy to increase marginal consumption is less likely to work in a world of globally aging populations. This is especially apropos to Europe, Japan, the U.S., and China (in just a short period they will have an older population that is larger than the entire population of the U.S.). We are beginning to see the end of the "shop till we drop" generation. And, the problem is not just with the big economies. Many other advanced smaller economies, such as Korea and Russia, are aging rapidly. Add these counties together and we have most of the world's GNP. There is a bit left in poorer parts of the world, however, for the most part the consuming part of the world is getting old.

We examine the demographic impacts that may disrupt traditional demand management policy tools, given the demographics of aging, the lifecycle of consumption for a country, as well as a person, with specific emphasis on household formation, and draw the implications. With demand management as a tool that is less useful – **forget the liquidity trap, it's the demographic trap.**

The paper first establishes demographic links that heretofore may be ignored when estimating the impacts of fiscal expansion. They are the direction of population growth in the major consuming nations, the aging of that population, and the implications for new household formation (a key to making demand management stimulus work).

Second, it considers the implication of aging with respect to the potential for significant household **deformation** in most of the consuming world.

Finally, it considers the systemic risk implications of household deformation on export led development, demand management policy, potential GDP growth, and asset valuations.

2. POPULATION TRENDS IN DEVELOPED COUNTRIES

Many studies, mainline business programs, and the media have all finally awakened to the implications of aging societies. The "graying of America" is a theme that has become popular. The implications for Social Security funding, Medicare funding, rising medical costs, are all now widely discussed. Simply do a search for graying of America and article upon article discussing the demographic trends we face are highlighted. Thus, these trends are known. And, to some extent it is understood that the U.S., with its demographic challenges, is young relative to many other countries (immigration has been a major contributor to keeping it younger). Moreover, we are beginning to see more mainline recognition that countries, as well as individuals, have a "lifecycle" with respect to the optimum age distribution for relative productivity.[1]

[1] Simply search on the internet for graying of America (http://bit.ly/2EQFADI) to get an idea of the amount of information available. With respect to the lifecycle discussion of a country see Silver and Wilford (2009) and Denby and Putnam (2017) for implications on productivity.

First consider the changing face of U.S. demographics. Figure 1 illustrates that the U.S. is moving from a young country (remember the 1960s) to a middle-aged country today. In 1960, 53% of the population was in the middle, older than 19 but younger than 65 years of age. In 2010 that number increased to 60% of the population, while the over 65 plus age group rose by 50% to almost 13% of the population. By 2050, projections suggest that the U S will be an old country. From a ratio of 19 and under to over 65 in 1960 of 4 to 1, the ratio has fallen to 2 to 1 by 2010 and is expected to fall much farther by 2050. More importantly, the middle group is the most productive. Yes, it will remain reasonably large but at the expense of the young, which are replaced by the old. As noted by Silver and Wilford (2009), it is the middle group that tends to have the relatively highest productivity (as well as savings and investments). Under 19 year olds are still consuming education, living off the society while not producing as a general statement. This fact tends to hold for over 65 as well, although the baskets of consumption goods tend to be very different. From a purely relative productivity perspective, one can observe that a country has a consumption-production lifecycle similar to that of the individual [Ando and Modigliani (1963)].

Germany and Japan were at their relative optimal demographic productivity (production focused ages relative to consumption ages) in the 1980s, while the U.S. was struggling to create jobs for a growing labor force. Political economics of the countries reflected these different realities. The U.S. needed to create net new jobs. In contrast, a steady state labor force existed in Germany and Japan, where similar numbers of individuals were leaving the labor force as were entering.

Today, as shown in Figures 2 and 3, the opposite situation exists, not from younger people entering the labor force in Germany and Japan, but by the intense graying of those countries, and indeed other European countries as well. This is laid out in Figure 4, which also includes data for Italy, one of the fastest aging European populations. In 1960, Italy and the U.S. had similar percentages of the population 65 and older with Japan younger still. By 2010 Germany, Japan, and Italy all had over 20% of their population over 65; the U.S. had creeped up, but only to 13%.

Figure 1: Changing face of U.S. demographics (1960-2010)

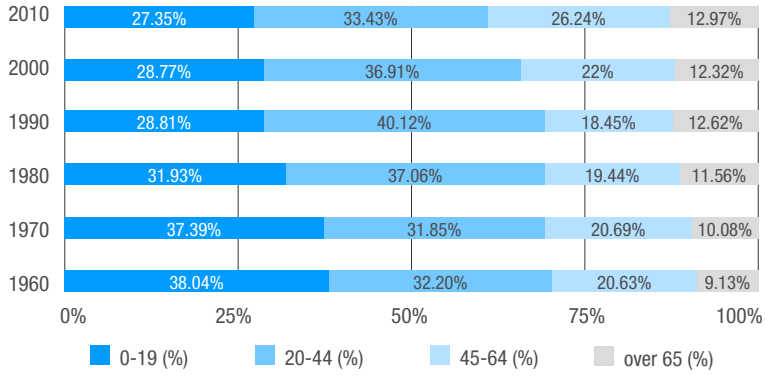


Figure 2: Changing face of German demographics (1960-2010)

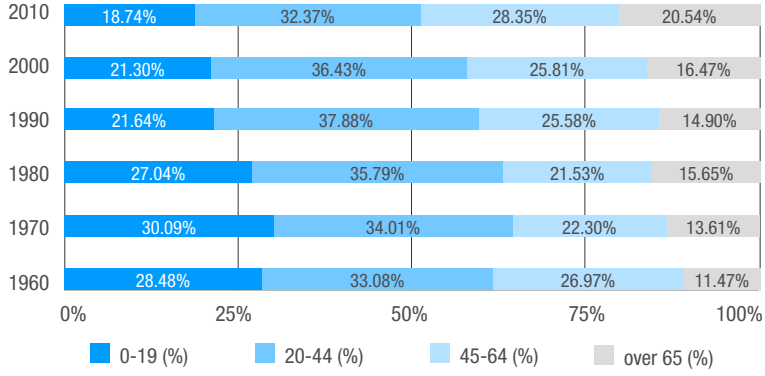
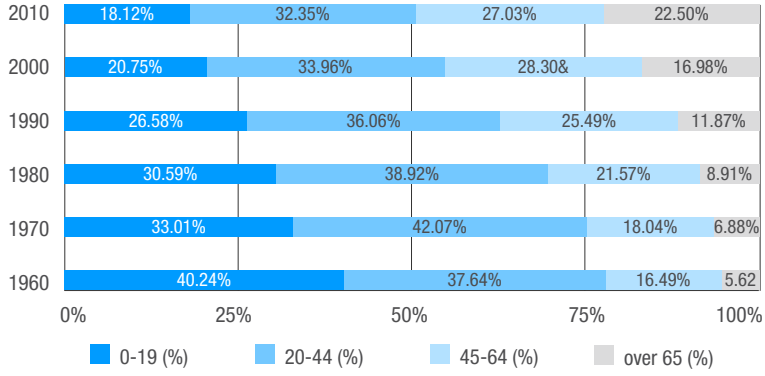


Figure 3: Changing face of Japanese demographics (1960-2010)



To grasp the extent of the problem, using U.N. forecasts, consider the same figures per above out to 2050 for Germany and Japan in Appendix 1.

Further, this is not just a European and Japanese phenomenon. Most middle-class OECD countries face this problem. The issue simply is when will it become a serious problem, not if. Figure 5 highlights this by looking at the percentage of the population of various high-income countries over 65 by the year 2050 (again, based upon U.N. projections).

China, although on net younger than the U.S. today, is expected to become "older" by 2040. As such, in just a short time, it faces similar demographic problems to those faced by Japan today. These demographic changes imply shifting consumption patterns, potentially lower productivity, as well as different political trends during the next decades.

Shifting demographics in the wealthy countries, the importing countries, have significant implications for development policies of those seeking to enter the OECD middle class club. Japan, South Korea, and China have used export led growth to propel themselves just as has Singapore and Hong Kong. Export led development requires other counties to be consumption driven. China is seen by many as the new consumer country, however, if it follows the pattern of demographic change forecasted, its older population may encounter the problems faced by Japanese during the 1990s, when the country entered a period of slow (often near negative) growth.

Figure 4: Percentage of population aged 65 years old and above

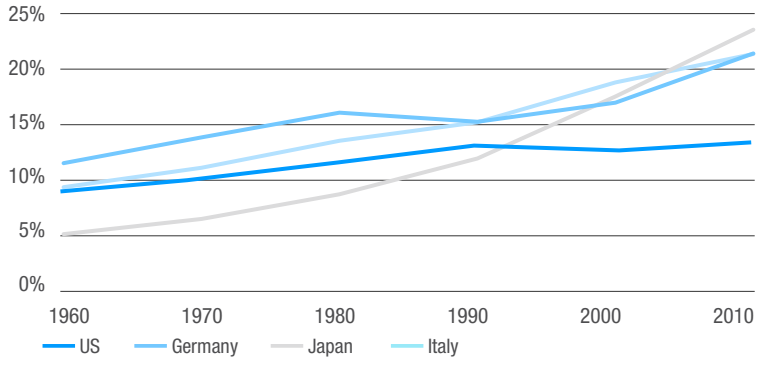
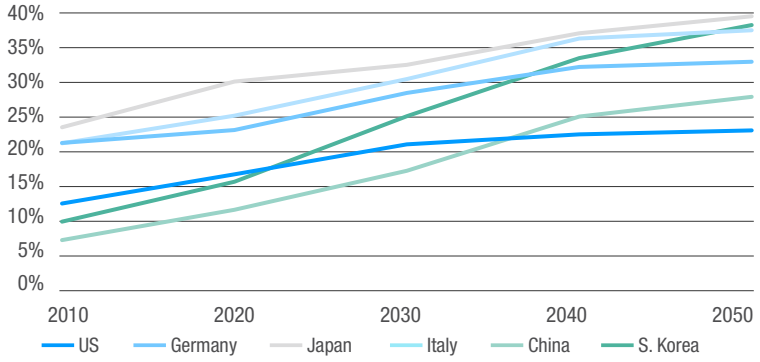


Figure 5: Percentage of population aged 65 years old and above: Projected



3. VEHICLE FOR CONSUMPTION-LED DEMAND – HOUSING AND HOUSEHOLD FORMATION

Since consumption-led demand can take many guises, it is impossible to choose one vehicle to model if we want to have an idea of demographic impacts on demand, as suggested above. In many cases, export-led growth occurs simultaneously with infrastructure and housing development. Separating the contributors to growth beyond simple accounting can be misleading. However, one indicator can be agreed upon as essential to generating demand: new demand for housing.

Housing demand may be driven by many factors. We can identify three distinctly different ones: (1) mass migration from the countryside as in China, (2) a growing population as in North America, and (3) rebuilding of housing after war (wars destroyed much of the housing stock in the latter half of the 20th century in Europe). Wealthy countries are not expected (hopefully) to be in a rebuilding situation resulting from war or a mass movement due to urbanization. It can be argued that the housing formation situation is stable and henceforth driven by real income and demographic factors. To the extent that China has entered the group of "wealthy nations," which could be a slight exaggeration, it will also experience a stabilization in demand for housing vis-à-vis the available population.[2]

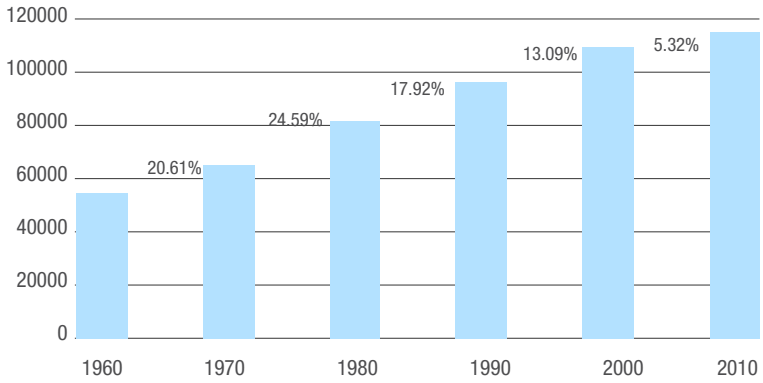
To examine housing demand, consider what drives the demand for new housing (not necessarily replacement) beyond the factors noted above. And then ask whether or not, one can segregate the factors driving that demand. To do so, one has to start with net household formation. Households form naturally as a population grows, thereby creating demand for housing. To examine a historical set of data one should consider household formation in the context of a stable (not disrupted due to war) environment.

Considering these factors, we choose to examine household formation in the U.S. in an attempt to determine the demographic (as well as wealth) factors that drive its formation. Tertiary conclusions can be drawn from Figure 6. Data are presented by decade. Household formation is slowing as the demographic makeup of the population is changing. It is obvious when the baby boomers enter the picture and create new families.[3] It is also obvious that household formation is now sharply slowing.

Household formation is essential for demand management policies, consumption driven, to work effectively. Formation of a new household implies building living quarters, buying goods to service the house, as well purchasing the items associated with creation of a household. Demand management tools, demand for exports from manufacturing countries, demand for consumption goods in general such as cars and garages, increase as households are formed. If no households are formed the nature of demand is different. If no net new houses are built how much copper for roofs and wiring is needed, for example.

At this juncture note that household formation is not the same as net new housing units created. Household formation is a primary factor determining demand for housing; however, the number of housing units built are also impacted by wealth characteristics. The second-home phenomenon is no doubt driven by the growth in wealth per capita over time as well as demographics (distribution of the age of a population as well as the general level of the population). For households, with exception of the extremely wealthy, to have multiple dwellings is a sign of the general level of wealth per household that many western countries have now achieved. To predict net housing demand, one would not only consider household formation growth but also wealth and/or per capita income growth. This paper focuses on household formation predictions. Based upon U.N. statistics one can reliably make

Figure 6: U.S. households and growth rates 1960-2010



[2] We are intentionally ignoring the massive underdeveloped populations in Southeast Asia and Africa, as well as many of the Latin American countries, which have young, non-urban, populations. Our focus here is on the developed economies in an attempt to isolate the demand conditions that support export-led growth for those countries.

[3] It is also obvious when their importance begins to disappear. Many, as noted by Denby and Putnam (2017), believe that one last hurrah is in the offing as the millennials finally begin to build households, however this may be a blip in the trend.

some predictions about household formation and thus the general implications for housing demand. It does not, however, address the second home phenomenon directly, albeit a consideration that must temper certain conclusions about the general impact of overall household formation (deformation) on the number of units demanded.[4]

3.1 MODELING HOUSEHOLD FORMATION IN THE U.S.

In general, household formation is driven by general population changes and by the ratios of the subgroups. Children are important in household formation, but they do not build the household; their adult parents do. Retirees on net tend not to create new households, but the opposite. Although to the extent that life expectancy increases are embedded in the ratios of the percentage of the population over 65, evaluating how rapidly households are deformed may not be as simple as one may think.[5]

One may create a model of household formation that is primarily defined by the overall population and what age groups tend to dominate the population.

A simple representation of a model to describe household formation can be characterized by the following equation.

$$HH = f(pop, e^{ratio\ under\ 19}, e^{ratio\ 19-44}, e^{ratio\ 44-65}, e^{ratio\ 65+})$$

Where HH is the number of households, pop is the level of the population, and the four ratios represent the distribution of the population.[6]

Pop and HH are levels and the demographic ratios are, by definition, already in percentages.[7]

The ratios as stated in the above model can misleading, however. A better way to think about the ratios and how they impact household formation is to consider the ratios that define the relative size of one group to another, not just the whole. The primary group that creates new households is no doubt the 19 – 44 age group. The next group likely to create a new household resides in the 44-65 age category (for example divorce may actually create a new household while remarriage undoes some of these temporary households).[8] The under 19 age group and retirees tend not to create net new households, while death or institutionalization may be necessary for the latter group to dissolve a household.

As such, we rewrite the ratios for focusing on the relationship of the under 19 years to the middle of the age distribution as well as the over 65 relationship.

This model can be rewritten as:

$$HH = f(pop, e^{age\ 19}, e^{age\ 65}),$$

Where age 19 = (% of over 19 to 65)/(% of under 19) and age 65 = (% of over 19 to 65)/(% of over 65)

The model in dlog terms can be rewritten as

$$PCHH = B_1\ PCpop + B_2 dage\ 19 + B_3 dage\ 65 + \epsilon$$

Where: PC is % change and dage is the first difference.

To estimate the model, we use annual data from 1960 through 2015. The demographic data is from the U.N. demographic database. Household formation data is supplied by the U.S. census bureau. All data are publically available.[9]

The results can be summarized as follows:

	COEFFICIENT	ST. ERROR	T – STATISTIC
B_1	1.041809935	0.165162383	6.30779186
B_2	0.198783216	0.145397565	1.367170186
B_3	0.804582496	0.262737264	3.062308268

The intercept is forced to zero.

Summary statistics are:

MULTIPLE R	0.900878997
R-SQUARE	0.811582966
ADJUSTED R-SQUARE	0.78458622
STANDARD ERROR	0.00750248

[4] Once household formation estimates are obtained one may consider this as a variable to model housing demand.

[5] We are focused on the U.S. at this point in the exercise, however, consideration of U.N. data for Russia provides a caution. Due to the declining longevity of the population, Russia is not expected to have as much of a growth in the above 65 portion of the population as in, say, Germany or the U.S. Shorter life expectancy or longer life expectancy are issues to consider in looking at the relevance of the ratios when comparing across countries.

[6] To repeat and highlight wealth is important for determining housing formation, less so for household formation. Underlying the data that will be deployed in estimating the model, however, is a basic assumption that the period covered for the U.S. starts with sufficient wealth to allow for smaller and smaller households. During most of the history of mankind multiple generations lived in the same household. With sufficient wealth, a household can become smaller as generations lived separately.

[7] Our modeling follows the typical of money demand equations that consider interest rates as a determining factor just as we consider ratios as determining factors.

[8] Paciorek (2015) describes multiple conditions that may impact household formation besides demographics.

[9] One criticism of this model could be that it does not have a wealth variable. Again, we are focused upon household formation not number of housing units in this formulation. However, in our estimations we did include it. The variable was insignificant as we would expect.

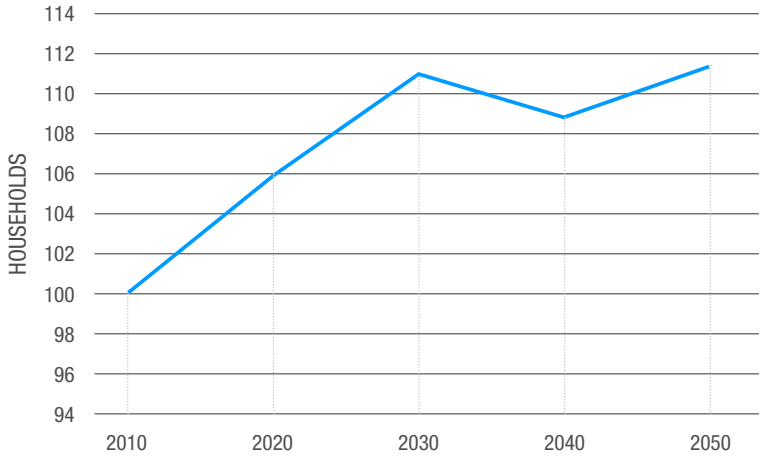
The fact that Coefficient 2 is insignificant is to be expected. More interestingly is the significance of Coefficient 3, supporting the notion that the aging process in reducing the ability for households to be formed (note that over 65 ratio is in the denominator) is more important than the number of those under 19. This may be because the key is the middle age group, from 20 – 65, which can create households rather than the number of children available to bring into the household.

Overall, the regression model appears to provide a reasonable explanation of household formation in the U.S. As such, we can utilize this simple model to forecast future household formation, given the demographic data from the U.N. Further, we will take the U.N. data for other countries and create a simple (recognizing its flaws) set of forecasts for the decades out to 2050 of household formation.[10]

4. FORECASTING BASED ON THE U.S. MODEL

Utilizing these estimates from the regressions above, we can build a forecasting model for U.S. household formation based upon the demographic makeup of the U.S. Figure 7 presents these forecasts. No doubt that these forecasts can be disputed on many grounds; however, the direction is clear. By 2040, net new household formation turns to deformation. Assuming a small bounce back toward 2050, household formation will have been very slow in thirty years. Moreover, the U.S. has (and will likely have) the best demographic outlook of all the countries analyzed.

Figure 7: U.S. predicted number of households (2010 base=100)



Although these estimates are for the U.S., and hence not easily extrapolated to the rest of the world, they may be useful in providing guidance to future net household formation. For Europe, North America, Japan, Korea, and other OECD countries that have already made the adjustment from a poor rural economy to an industrial or post-industrial urbanized economy, we may use the U.S. model as a baseline to forecast net new household formation. It could be argued that near term movements of populations from poorer rural areas of China to urban industrial cities obviates the usefulness of the any model based on U.S. data. Longer term, however, there is every reason to assume that the general conclusions will hold for China as well, once the urbanization process is completed.

[10] We are aware of the shortcomings of using U.S. data to forecast household formation for other countries. Since we are interested in gross implications consider the results with a wide area of possible deviation.

Table 1: Household formation projections by country – 2050

DECADE	COUNTRY	PERCENTAGE CHANGE IN NUMBER OF HOUSEHOLDS
2010	China	-33.861%
2020		-12.869%
2030		-15.240%
2040		-4.294%
2010	Germany	-6.388%
2020		-7.569%
2030		-9.452%
2040		-3.254%
2010	Japan	-17.747%
2020		-12.428%
2030		-6.852%
2040		-6.685%
2010	Korea	-17.264%
2020		-17.710%
2030		-13.120%
2040		-6.632%
2010	Italy	-17.359%
2020		-4.913%
2030		-3.384%
2040		-5.316%
2010	U.S.	5.924%
2020		4.843%
2030		-1.972%
2040		2.359%

Table 1 presents the projections for Japan, Germany, France, Italy, Korea, and the U.S. Clearly these projections will not be correct. The input factors vary widely depending upon the assumptions in the U.N. forecasts and the errors can be quite large.[11] This aside, the implications for the direction of household formations are observable. In each of the above cases, it is forecasted that movement is towards household deformation, not net positive new formations, with, as expected, Japan leading the way. Percentage changes represent 10-year household formation.

Since China is the second largest economy and expected to become the world's largest by many, a similar analysis for the Chinese economy is made. In this case we assume that the rural to urban process is completed by 2040. As such, the forecasts for the periods 2010, 2020, and 2030 may not be good forecasters of overall housing demand. Still, the implications of this are startling. It may be that the housing boom will be over much sooner, suggesting a potential housing market collapse, given the projections of household deformation now occurring.[12]

[11] The U.N. database for demographics is the source. Many variants exist for possible demographic movements. We have chosen to work with one that is moderate, since it neither assumes a steady state of birth and death rates nor does it assume a return to a more fertile birth trend. Needless to say, deviations in immigration from historical norm will impact the actual outcomes as well.

[12] An interesting aside is that many are now arguing that China has overbuilt already given the number of structures that are not occupied. If this argument is correct, then a great deal of assets now on bank balance sheets are actually worth much less than face value. This was pointed out by Christopher Rapcewicz to one of the authors as a potential issue that could plague the financial system of China.

Only the U.S. seems to avoid severe problems, with only one decade of suggested deformation. Moreover, the small size of this projected dip suggests that it is possible that the U.S. can avoid deformation completely. Clearly, however, to believe that the household formation impact on demand will resemble anything like that of the past 50 years would be inappropriate.

In contrast to the U.S. observe forecasts for Italy. Anecdotally, towns are already being emptied of people. As such, the forecast for 2020 does not seem out of touch with reality. We are already observing the phenomenon. And in the case of Japan, the size of the negative numbers are startling. These are percentage changes per decade. Projected population declines and the aging of the population are all consistent with today's observations. If these projections are close to reality, assuming a decline in households of 30 to 50% over the next 40 years is likely. If so, one can then extrapolate the implications for the excess supply of housing arising from this trend. For South Korea the situation is just as bad or worse. And, for Germany the trend is clearly similar.

Household formation in the wealthy countries, with the exception of the U.S., is clearly headed toward significant deformation. Household deformation is a new phenomenon for modern economies. Due to population declines during periods of famine and disease during earlier centuries we know that the implications for growth (stagnation) are many, including significant political upheavals. As such, it behooves us to understand some of the implications now, so that policy decisions will reflect these realities before significant decisions (errors) are made. Moreover, the implication for the banking and insurance industries are enormous given that housing is a crucial part of their business, either directly or indirectly through asset accumulation via the capital markets.

"It is the demographic trap — forget the liquidity trap — that will negate historically observed positive aspects of demand management policy."

5. POLICY IMPLICATIONS OF HOUSEHOLD DEFORMATION

Most vulnerable to the implications of the aging of the populations is the Keynesian multiplier methodology engrained in most economic growth models. As such, standard pump-priming policies, shovel-ready stimulus may not work. It will not drive demand in the same manner as it did in the 60s or even later decades. And, with potential household deformation such policies may simply be distortionary, without any benefit.

Further, it may be difficult for export-led economic growth to work. Trade has always been a vehicle for greater economic well-being. It has been the source of much of the ending of poverty for many. Japan and Korea used it to develop, just as China has done so more recently. However, this tool will be under threat as a vehicle for growth if the wealthy nations are not growing households, demanding the types of goods that go with such growth.

Although policy may be the most important issue raised by household deformation, there are implications for the capital markets as well. As household deformation becomes evident, wealth captured in existing housing is reduced due to excess supply. As such, housing wealth in many countries may already be overestimated, raising issues for pension funds, banks, and insurance companies that depend upon MBS portfolios. Much as new solar technology may make oil and gas in the ground less valuable as some argue, banks and the capital markets may already be overestimating the value of housing held in their portfolios, just as in 2008 -09. Only this time the overhang may be unending or at least for the life of many securities now held by those firms. This creates a large systemic risk for the global financial system.

The world economies rebounded from the last crisis – it is argued by many though disputed by some – via government deficit (stimulus) policy. Moreover, this fiscal response (evidenced by government debt levels) was supported by zero interest rate policies and the socialization of many of the world's banking risks. If the next systemic financial crisis arrives during a period of extreme household deformation, these policies may not be sufficient or even useful. Indeed, if the capital value decline affecting the financial system is driven by household deformation, then fiscal stimulus policies will not likely have a positive impact, thereby leaving the financial system in a precarious position. Traditional models of behavior cannot be depended upon, implying traditional models cannot be depended upon. History should not be taken as a guide to the future.

Monetary policy predictability becomes more important; expansionary policy to stimulate demand may not work with lags as in the past. Indeed, if the policy works at all as intended, lag effects may be unpredictable, especially with a globally integrated economy. Unintended consequences may lead to political disquiet as wealth is transferred from savers to borrowers, without enhanced wealth creation for the middle and lower income earners (leading to a skewing of income distributions due to increased return on leveraged capital with low interest rates). At a very least, such policies will create distortions that are different from the ones that may have been created during the 30s or 50s with such policies.

What then can governments do? If they cannot tax and spend to create a multiplier effect; if central banks cannot push us out of the demographic trap is there any solution? If the benefits of export-led development models that helped Japan, Korea, and China to develop may not work in the future, is there anything left?

It is the demographic trap – forget the liquidity trap – that will negate historically observed positive aspects of demand management policies. There is little that central planners can do about it other than accept that their top-down stimulating policies will not work in time, that their forecasts are not likely to be anywhere close to correct, and that slower growth and low inflation are on the cards unless a new policy direction, not rehashed neo-Keynesian prescriptions, are found.

APPENDIX 1

Figure A1: Germany demographic changes (2010-2050)

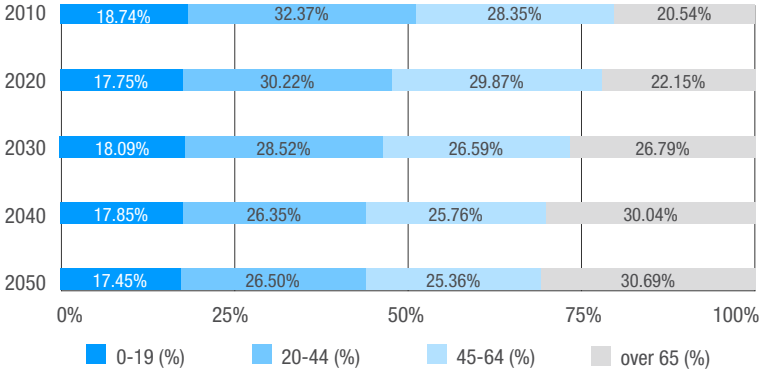
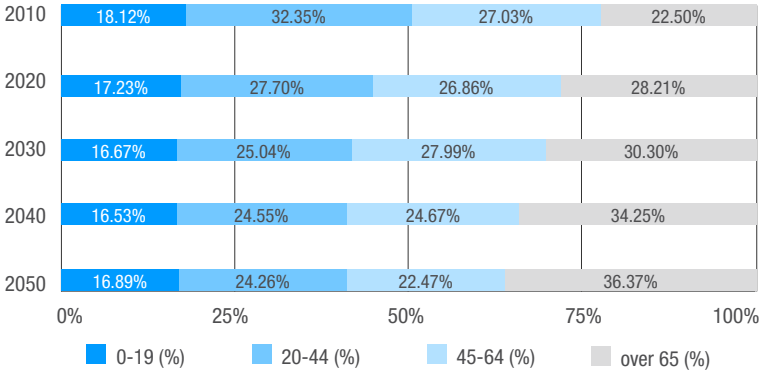


Figure A2: Japanese demographic changes (2010-2050)



References

Ando, A., and F. Modigliani, 1963, "The "lifecycle" hypothesis of saving: aggregate implications," American Economic Review 53:1, 55-84

Bloom, D., 2016, "Demographic upheaval," Finance & Development

Denby, N., and B. H. Putnam, 2017, "Sluggish wage and productivity growth: structural changes are to blame," The Hedge Fund Journal, 126, Fall 2017, London

Dunne, T., 2012, "Household formation and the great recession," Economic Commentary, number 2012-12

Ermisch, J., 1999, "Prices, parents, and young people's household formation," Journal of Urban Economics 45:1, 47–71

Masnick, G., D. McCue, and E. Belsky, 2010, "Updated 2010–2020 household and new home demand projections," Joint Center for Housing Studies, September

Paciorek, A., 2015, "The long and short of household formation," Real Estate Economics 44, 7 - 40

Silver, S. J., and D. S. Wilford, 2009, "The impact of demographics on economic policy: a huge risk often ignored," Journal of Financial Transformation 25,

Yardeni, E., and M. Quintana, 2017, "US demography: household formation, homeowners, and renters." http://bit.ly/2EPS8LE

CURRENCY

PREVIOUS EDITIONS OF THE CAPCO JOURNAL OF FINANCIAL TRANSFORMATION ARE AVAILABLE AT **WWW.CAPCO.COM/INSTITUTE**

- Security and identity challenges in cryptotechnologies
- Economic simulation of cryptocurrencies
- Narrow banks and fiat-backed digital coins
- Quantitative investing and the limits of (deep) learning from financial data

Security and identity challenges in cryptotechnologies[1]

JOSÉ VICENTE | Chairman of the Euro Banking Association's Cryptotechnologies Working Group
THOMAS EGNER | Secretary General, Euro Banking Association (EBA), on behalf of the working group

ABSTRACT

The use of cryptotechnologies (CTs) in transaction banking is currently widely discussed in the financial services industry. Since the description and publication of the Bitcoin system in 2008, the potential of CTs (also known as distributed ledger technology) to simplify and enhance traditional processes in transaction banking has been attracting much industry attention and debate. Use cases have been defined and discarded in the search for implementations that would increase efficiencies and/or unlock new business opportunities for both financial service providers and their customers.

In 2015, the Cryptotechnologies Working Group of the Euro Banking Association (EBA) started to explore the practical implications, opportunities, and challenges of CTs in transaction banking. Composed of payment practitioners from banks across Europe, the working group has been looking into concrete use and potential business cases, e.g., foreign exchange (FX), real-time payments, trade finance, or international payments.

For its current publication, the working group examined the use of CTs in processes where data security and integrity are key. The resulting paper, which is reprinted here in form of an article, covers two use cases – third–party authorization (both from a bank and customer perspective) and know your customer (KYC) and due diligence processes. It describes how banks, as well as their customers and other stakeholders, can experience the benefits in terms of transparency, speed, and efficiency that the use of CTs can offer in these contexts without having to compromise on data security and regulatory compliance.

1. INTRODUCTION

Data security is of paramount importance in financial services. The secure storage and exchange of information is one of the key services banks offer their customers. As end-user demands evolve and new regulations, such as the second Payment Services Directive (PSD2), mandate more open exchange of data, financial institutions have been exploring the possible roles of cryptotechnologies (CTs)[2] in this changing environment. While rules regarding access, speed, and participation are changing, ensuring the integrity and security of financial data will continue to be a necessity.

The focus of this article is on how CTs can maintain or improve data security and integrity while opening new opportunities for financial institutions. CTs can help financial institutions to both enable regulatory compliance and improve service to end-users while lowering costs and providing future flexibility as payments and financial services continue to evolve. While the full value of CTs will come with widespread usage, many banks today are pursuing an incremental approach to adoption. This approach involves an assessment of how CTs interact with legacy systems to determine where distributed ledger technology (DLT) fits in an institution's technology stack. The use of CTs can occur within a single organization, an entire payments community, cross-domain, or even across borders.

Through numerous group discussions and demos from banks and software providers, the Euro Banking Association's Cryptotechnologies Working Group has analyzed how CTs may help achieve higher efficiencies by improving speed, accessibility, and operability to facilitate new services in an environment marked by new commercial and regulatory developments regarding access and control of data. Two use cases were examined, covering third-party authorization (both from a bank and customer perspective) and know your customer (KYC) and due diligence processes. Financial institutions are already exploring the use of the technology in these areas, and have been developing fit-for-purpose DLT solutions. The challenge for financial institutions in adopting CTs will be in re-thinking their implications on existing IT and business processes while maintaining flexibility and adaptability for future needs.

This article will begin with an explanation of the characteristics of evolving DLT solutions. It will then examine two use cases related to third-party consent management and sharing of KYC attributes within and between organizations, including the benefits and challenges associated with each use case. It will end with a look ahead at how financial institutions can benefit from increased industry adoption of CTs.

2. DLT CHARACTERISTICS

Previous reports of the cryptotechnologies working group[3] have identified four key aspects shared by various CT solutions:

1. A shared, uniform ledger that is replicated among all participants over a network of interconnected computers.
2. Security and accuracy of the ledger is ensured through cryptographic methods.
3. Control of the ledger is decentralized among network participants (no single central authority).
4. Once verified, transactions on the ledger are fixed and indisputable.

CTs were initially designed to ensure finality and transparency of transactions across a distributed network. These core features were not developed with legacy bank processes and financial regulations in mind. With an increasing number of financial institutions actively exploring the use of CTs, there have been several important developments in DLT solutions designed to help the technology adapt to the business, legal, and regulatory realities of financial organizations. Financial institutions using CTs today must make determinations on a few additional key aspects that can affect data security.

2.1 Permissioned ledgers and limiting access

Early implementations of CTs, such as Bitcoin, were unpermissioned (and continue to be so), meaning that any party can join the network and verify transactions. In the traditional, highly regulated payment infrastructure business, on the other hand, access to messaging and payment networks is always permissioned. This is not expected to change with the use of CTs. Permissioned ledgers allow more control over who has access to the ledger and which role is assigned to each participant.

[2] Cryptotechnologies are also referred to as distributed ledger technology or DLT. The term "blockchain" will not be used in this report, as it is a specific type of distributed ledger and the focus of this report is on the technology in general.

[3] EBA, 2016, "Applying cryptotechnologies to trade finance," Euro Banking Association, May, http://bit.ly/2l87LPR; EBA, 2017, "Cryptotechnologies in international payments," Euro Banking Association, March, http://bit.ly/2qCfChP

Today, central authorities, such as national central banks or other market infrastructure providers, play the role of maintaining and verifying ledgers, but with DLT, this role can be divided over multiple entities in the network. Having unauthorized entities involved in verifying new transactions would, however, be too risky for financial institutions and their customers. Thus, while control can be decentralized, it will still have to be exercised by authorized parties. When using CTs, all entities involved in verifying new ledgers must, therefore, be authorized.

Authorization to view information on the DLT ledger will also be controlled. Initiatives such as Multichain,[4] Ripple Connect,[5] and Hyperledger Fabric[6] all offer permissioned access to view ledgers, ensuring that all nodes can be identified and are authorized to access information on the ledger. These entities can then be given permission to access information on a need-to-know basis. Permissioned access to the ledger will be vital for creating the trust needed for institutions to exchange information between organizations and across borders. These layers of access ensure that all participants in a ledger meet certain standards for verifying information and/or accessing information, helping maintain data security in the network.

2. PRIVACY OF INFORMATION

While CT solutions employ various methods to ensure confidentiality for participants on the ledger as data is shared across the network (e.g., by using pseudonyms for each party sending and receiving information), the amount of information shared on the ledger does leave open the possibility of reverse engineering transactions to determine which banks or bank customers are directly involved in a transaction. This has understandably led to concerns among institutions for whom confidentiality is paramount. To combat this, some CT initiatives have developed private ledgers that ensure that information exchanged as part of a transaction is only visible to the parties involved in that transaction. The Corda platform by R3[7] is a prominent example of a private ledger developed with involvement from leading banks around the world. The ability to exchange information privately on a cryptotechnology platform may be a key enabler of widespread adoption going forward and allow experimentation without the risk of disclosing sensitive information of any kind.

Banks using DLT must determine which information is most suited to be exchanged internally or externally using CTs. For more sensitive information, participants should choose which information is kept on-ledger (using DLT) and which is stored off-ledger (using traditional systems like databases or data warehouses). This will necessarily involve an analysis of whether the cost of segregating data between ledgers outweighs the cost-savings and increased efficiency that can come from using CTs.

2.3 Immutability of data

Financial services data is always subject to change, particularly data related to a customer's identity. Financial institutions thus need to be able to amend or withdraw data as information evolves or regulation (or a user) demands. In other words, certain data needs to be revocable. The need for revocability was not a key concern in the first generation of CTs. Banks have worked together to develop new solutions to this problem by using private channels within a CT solution or by holding sensitive data off-ledger and using a distributed ledger to exchange specific attributes. These developments seek to overcome concerns related to commercial sensitivity of data.[8] When determining how to use CTs, financial institutions must consider data immutability to select the use cases and approaches that are most appropriate for a distributed ledger solution.

2.4 Participating nodes on the ledger

Determining which entities can participate as nodes on a CT ledger will be a key issue for banks. A ledger used internally by a bank may be made up of individual nodes that represent entire departments, or individuals within specific departments. Ledgers used across organizations (for instance, between a bank and its domestic or foreign subsidiaries) may see each node representing an entire organization, departments within each organization, or individual employees. Banks need to determine which actors or entities need direct access to the ledger to ensure proper representation and avoid bottlenecks while protecting access to sensitive customer data.

[4] https://www.multichain.com
[5] https://ripple.com
[6] https://www.hyperledger.org/projects/fabric
[7] https://www.r3.com and https://www.corda.net
[8] Scalability is an additional concern in this space. The more data stored on the ledger, the bigger each copy of the ledger will be as it is shared among all nodes. Having some data stored privately or off-ledger means that each copy of the ledger held by all nodes will be smaller, thus increasing overall scalability.

As new regulations, such as the General Data Protection Regulation (GDPR),[9] open up space for end-users to control their own data, it is possible that consumers and businesses could eventually represent nodes on a CT ledger. This is unlikely to occur in the near future, but banks should start thinking about possible implications, such as how to enable enhanced user control without opening up access to a distributed ledger directly, e.g., via "application programming interfaces" (APIs).

3. THIRD-PARTY PROVIDER (TPP) AUTHORIZATION AND CONSENT MANAGEMENT

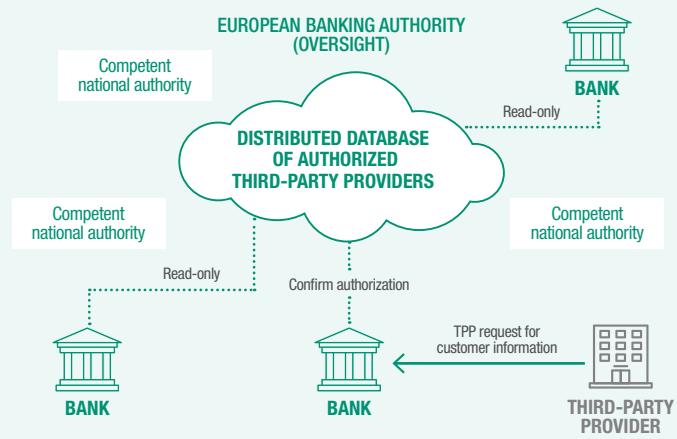
With new regulations, such as the revised Payment Services Directive (PSD2) and the GDPR due to become applicable in 2018, European banks and third-parties will need to undergo a shift in how they manage consent for financial services. Banks will be required to provide access to payment accounts upon their customers' requests, while having to ensure at the same time that end-users and third-parties are properly authorized and permissioned to access data. This will require a change in business practices that will be aided by the widespread adoption of technologies such as APIs. CTs also hold the potential to help banks comply with these new regulations while preparing for a future where the controlled sharing of data and value within and between organizations is facilitated on a large scale.

CTs can enable enhanced consent management in two ways: by giving European banks an up-to-the-second, unified view of all authorized third-party providers in Europe and by giving end-users control over which entities they have authorized to access their bank account information. Each of these solutions addresses a different side of the same problem by ensuring that third-party access to bank account data is authorized and that end-users retain control of their bank account data.

3.1 A uniform view of all authorized TPPs for banks

The bank side of consent management involves providing a list of all authorized third-party providers to every bank in Europe. This would allow European banks to instantly check to see if a third-party requesting access to a customer's bank account data is authorized to access that data under the PSD2. In theory, CT solutions are not necessary for performing the task of consistently updating a ledger of authorized entities that can be used by banks throughout Europe. Indeed,

Figure 1: Managing third-party provider authorisation for European banks



the European Banking Authority released a consultation paper in July 2017 proposing "a technological solution that will support both manual insertion and automated transmission of information by competent authorities (CAs) to the European Banking Authority (EBA)."[10] But in practice, having a single central authority update a ledger entails several issues, including: (1) determining which entity updates the ledger; (2) whether a single entity is needed for all of Europe if authorization is coordinated between national authorities; (3) ensuring that all banks across Europe have a uniform copy of the authorization ledger that can be updated in real time and that experiences little downtime in availability (the automated solution outlined in the EBA Consultation paper discusses updating information from national CAs on a D+1 basis); and (4) avoiding errors and omissions that can occur when banks manually check a routing table for information.

Using DLT to manage and check data on authorized third-party providers could enable a more efficient, cost-effective, and reliable authorization process.

A CT authorization ledger would ensure that all European banks would have a shared, single view of all authorized third-party providers in Europe. Under the PSD2, the European Banking Authority is mandated to establish and maintain a central register of third-party

[9] http://www.euGDPR.org
[10] EBA, 2017, "Consultation paper on the draft RTS and ITS on the EBA Register," European Banking Authority, http://bit.ly/2G0a72x

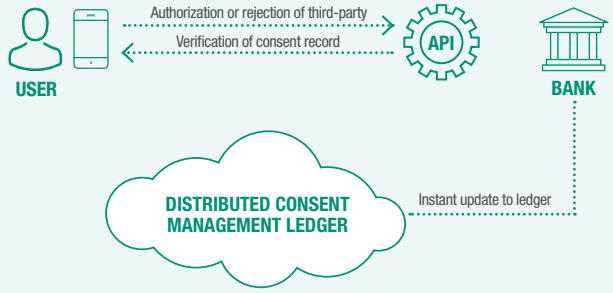
providers as authorized by CAs in E.U. member states. European banks will rely on this register to verify that third-party providers (both "payment initiation service providers" (PISPs) and "account information service providers" (AISPs)) are entitled to provide payment services to end-users under the PSD2. With so many participants involved in updating, managing, and using the register, the use of DLT could enable greater speed and efficiency, with lower cost and a lower risk of unauthorized access to bank account information. With a distributed register for third-party providers, all national CAs could instantly update the ledger under rules set by the European Banking Authority without the need for manual interventions. This includes a record of a TPP's authorization under the PSD2 as well as a record of exactly when a TPP loses that authorization for any reason. European banks could then have read-only access to this ledger to verify any TPP requesting bank account information of one of the bank's customers. Once the entity is verified as being an authorized TPP under the PSD2, it would only have to provide proof of a customer mandate to receive specified access to customer information. Should a TPP lose its authorization for any reason, the automated verification using CTs could also void all existing consent given to the TPP from end-users.

Banks will need to update internal IT and business processes to accommodate this, but this process is already under way with the development of APIs and the move to faster payments. Scalability concerns would not be an issue as banks would merely be accessing the crypto register to verify the data stored on the ledger; they would not need to add data or transactions to the ledger itself. The implementation of PSD2 will necessitate deeper coordination between banks throughout Europe and a harmonized process for authorizing third-party providers to access customer data. DLT can play an important role in this process by enabling banks to instantly check third-party authorization and ensure that unauthorized entities do not gain access to sensitive bank account information.

3.2 Customer consent management using DLT

CTs can also enable enhanced control of third-party provider authorization for end-users. This is particularly vital considering the PSD2, which became applicable throughout Europe in January 2018. Under the PSD2, consumers and businesses will be able to authorize third-parties to access their bank account data for information or payment initiation services. This use case will also have relevance to the GDPR, which will apply from 25th May 2018. The GDPR will also require explicit consent from end-users for the processing or sharing of certain customer data, as well as a "right to be forgotten." In the near-term, consumers and businesses may not be given full control to manage all of their digital data across numerous platforms as direct participants in a DLT ledger. Banks will have to ease this complexity by providing their customers with interfaces to help control and manage data, and a fully auditable record of which entities have been given consent to access bank account data could be a key enabler of regulatory compliance and an improved customer experience.

Figure 2: Distributed consent management ledger



Currently, many banks lack comprehensive and well-integrated end-user consent management systems. CTs can provide the needed technology for developing such systems. Being greenfield implementations, CTs offer a compelling case for implementing consent management systems, with integration to legacy systems and processes occurring via APIs. A CT ledger can give a bank a single, unified view of which permissions their customers have given to various third-parties or divisions within the bank without the need to store sensitive data itself on the ledger.

Bank customers may not have direct access to a CT ledger, and banks will play a crucial role in providing straight-forward and user-friendly interfaces to enable advanced functionality for end-users. Users could either give or withdraw consent for third-parties to access their bank account data via a front-end app on a mobile phone or online. Banks would receive these requests via APIs and then immediately (and immutably) store the record of consent on the DLT ledger. Once consent

is revoked, the record on the ledger would be instantly updated to reflect this. Any future disputes could easily be resolved by reviewing the record of consent on the ledger, and users could verify their information on the ledger via the interface provided by their bank. The only information stored on the ledger would be the record of consent given to each third-party; the end user's bank account information would be stored off-ledger at the bank as it is done today. End-users would be able to review this record online or via a mobile app, giving them added control and security of their data even when using multiple third-party apps or bank products.

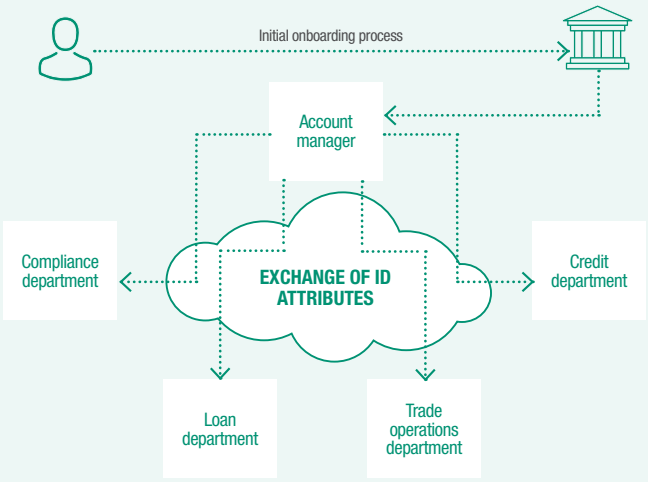
The use of DLT for customer-facing consent management would help comply with regulations such as the PSD2 and the GDPR, while also providing a frictionless experience for bank customers. Permissioned ledgers help minimize scalability concerns, and occasional updates to the record of consent do not represent a high volume of transactions. The concern about immutability of data on a ledger would not be relevant because no private customer information is stored on the ledger, only a record of when consent is given and taken away. In fact, the immutability of this data would be a positive aspect due to the ability to fully audit an entire history of customer authorizations.

3.3 Benefits and practical considerations for consent management

Benefits of CTs for consent management are: (1) greater speed and efficiency in ensuring TPP authorization and customer consent; (2) improved customer experience through enhanced control over third-party access to data; (3) aids compliance with regulations such as the PSD2 and GDPR; (4) instant identification of authorized TPPs increases efficiency and lowers risk of fraud or error; and (5) increased transparency due to fully auditable record of all entities that have been authorized under PSD2 or given access by customers to bank account information.

Practical considerations and challenges include: (1) determine where CTs fit in IT stack and update business and data governance processes accordingly; (2) analyze costs of segregating data between ledgers; (3) develop interfaces allowing customers to interact with DLT ledger and use APIs to automate this process; and determine which entities are represented as nodes on ledger and what type of access each participant should have (read-only, verify new ledgers, etc.).

Figure 3: Distributed identity attribute management



4. EXCHANGE OF KYC INFORMATION WITHIN BANKS AND WITH SUBSIDIARIES

The complexity and redundancy involved in KYC processes today is a big driver of cost for banks and their customers. As was explored in the March 2017 report "Cryptotechnologies in international payments"[11] published by the Euro Banking Association's Cryptotechnologies Working Group, DLT offers huge opportunities to lower cost in the complex value chain of international payments. But CTs also offer banks a way to rationalize their internal onboarding processes and complex records of identity for a single customer. Further benefits could be achieved by opening access to identity information between a bank and its own subsidiaries. CTs can provide a single internal source of truth on a customer's identity, which could help reduce the time it takes to onboard clients and avoid potential complexities and errors that result from a fragmented information onboarding process.

There are two aspects of KYC that banks must consider: Customer due diligence related to onboarding a client and the anti-money laundering (AML) screening of a payment itself. This use case will deal with the former. By facilitating secure access to KYC information between multiple parties within banks or banks and their subsidiaries, DLT can reduce costs and onboarding time. It can also provide opportunities for banks and

[11] EBA, 2017, "Cryptotechnologies in international payments," Euro Banking Association, March, http://bit.ly/2qCfChP

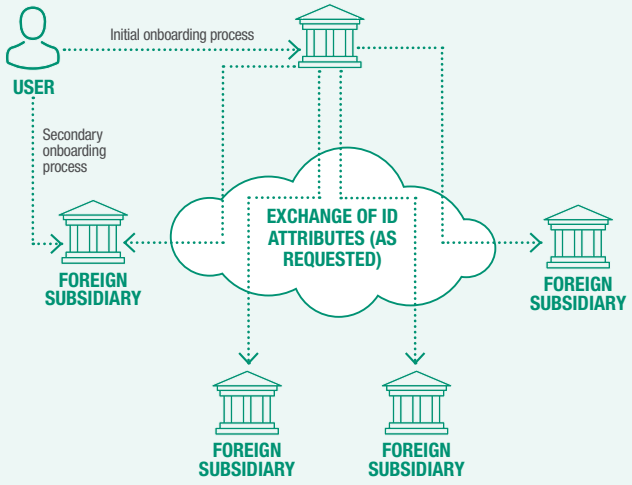
their subsidiaries to offer products tailored to the needs of their customers with few redundant processes.

4.1 Distributed identity attributed management within banks

The sharing of customer KYC information within banks can be a very fragmented process (if it happens at all) that is marked by redundant procedures across various divisions. Frequently, a customer must resubmit identity information or documentation when applying to use a new product or service with their bank. Banks find it difficult to share information internally mainly because of their highly paper-based processes and a build-up of internal silos between divisions after bank mergers. With a highly paper-based onboarding process and the complexity of communication between siloes, it is often easier for the bank to perform redundant onboarding processes with existing customers than to rework internal processes to enable the fast and open exchange of customer information. Many bank customers do not understand why they are required to resubmit information that has already been shared with the bank, which leads to a poor customer experience.

CTs can help facilitate the reuse of customer data within banks. KYC and onboarding processes would remain largely the same, but the storing and exchange of this data would be much more efficient, secure, and faster. After a division within a bank onboards a customer, the KYC information obtained could be stored on a backend system used within the bank. The account manager could then use this information to build a customer's identity that consists of the attributes obtained in the onboarding process and supplemented with additional information as the customer makes transactions. When a customer requests a product or service from another division within the bank, that division can call on the customer's internal identity profile to request access to attributes needed for the additional service. These attributes would then be exchanged internally via DLT, enabling that division to provide the service to the customer without asking for any additional information (or, if additional information is needed, the customer only needs to provide specific information instead of resubmitting prior documentation). There would be no need to share entire documents – only the specific attributes needed for the request would be shared within bank divisions.

Figure 4: Exchange of ID attributes with subsidiaries



4.2 Distributing identity attributes between banks and their subsidiaries

This principle could be expanded beyond the (local) confines of a bank as well. Some bank customers, particularly corporates, do business in many jurisdictions. Banks with subsidiaries in multiple markets often leverage their size to attract corporate customers with diverse needs for payments and other financial services. The fact that subsidiaries are often located in separate jurisdictions means that corporates need to go through entirely new onboarding processes each time they use a product or service from a local subsidiary. Despite being the same bank, or a subsidiary providing ancillary services necessary for global commerce (such as insurance), corporate customers still have to go through the complex process of providing necessary documentation and KYC information. Banks can leverage DLT to allow for a more seamless customer experience across jurisdictions while maintaining security of information and compliance with regulations in multiple markets.

In the cross-border space, banks will have to ensure that the exchange of identity attributes across jurisdictions does not compromise compliance with local laws and regulations. Banks may consider using smart contracts to ensure that only valid and legal identity attributes are shared with subsidiaries abroad. Any restrictions on sharing an attribute or piece of data in any jurisdiction can be embedded in the smart contract code to ensure that banks comply with local regulations without the

need to rely solely on manual processes. In some jurisdictions, the use of smart contracts may not be enough. Countries such as the Netherlands currently mandate that the entity that onboards a customer remains liable for conducting KYC checks accurately. As regulations such as the GDPR seek to give users more control over their data and technology makes the concept of self-sovereign electronic identity feasible, these regulations may need to be revisited to ensure that they are fit for purpose.

The absence of a global identifier in correspondent banking is a major hurdle for banks today. Although CTs can enhance efficiency and speed while enabling the secure exchange of identity attributes (as opposed to full Id documents), the banking industry still needs to harmonize various approaches to legal identifiers and determine whether existing initiatives such as "legal entity identifier" (LEI) are adequate or whether additional solutions are needed. The lack of a global market practice for exchanging KYC information is another challenge to sharing identity attributes across borders. Different markets require different information for customer onboarding and screening. The higher due diligence requirements needed in correspondent banking mean that any CT solution aimed at exchanging KYC information will have to account for requirements in different jurisdictions and maintain flexibility to deal with regulatory changes as they occur.

4.3 Benefits and practical considerations for KYC management

Benefits of CTs for KYC management include: (1) increased efficiency in exchanging information between bank departments and between banks and their subsidiaries; (2) maintaining data security and compliance with regulations allows banks to shift focus to improving customer experience and attracting new users; (3) having information that is machine-readable can reduce error rates and improve speed; and (4) potential for cross-selling of products to consumers and businesses based on identity profile.

Practical considerations and challenges include: (1) lack of global market practice for exchanging KYC information and absence of global identifier for financial services will continue to be hurdle in correspondent banking; (2) data protection laws in some jurisdictions prohibit the exchange of certain data between institutions (analyze which jurisdictions are most attractive for DLT solutions); and (3) need to ensure revocability of data in line with data protection requirements.

5. LOOKING AHEAD

The use cases examined in this article can help banks as they deal with evolving customer demands and new regulations calling for faster information exchange and greater transparency in financial services. An incremental approach to DLT adoption gives banks the opportunity to assess how the technology interacts with existing internal systems and interbank networks and to examine new use cases that can increase efficiency, lower costs, enable new products, and improve service for their customers. As some members of the EBA's Cryptotechnologies Working Group have reported, internal proofs of concept with CTs have also helped trigger a wider conversation about the role their organizations play in providing payment services to their customers and where they fit in the payments value chain going forward. This fundamental assessment of the role of banks is vital at a time when new industry players are entering payments and new regulations demand that banks rethink their role as a one-stop shop for payments and banking services.

While the gradual adoption of CTs can bring tangible benefits to banks and other players in the short-term, the full benefits of DLT will not be unlocked until the technology is used by a wide variety of financial industry stakeholders. Industry collaboration will be key. Banks should work closely with other financial institutions and regulators to explore the effects CTs have on data security, processing efficiency, regulatory compliance, and customer experience. CTs can help open new horizons on how to explore solutions to existing problems. As DLT adoption evolves from internal use cases to include multiple organizations in multiple jurisdictions, financial industry stakeholders and their customers will experience the full value of transparency, speed, and efficiency without the need to compromise on data security and regulatory compliance.

Economic simulation of cryptocurrencies[1]

MICHAEL R. MAINELLI | Chairman, Z/Yen Group, UK and Emeritus Professor of Commerce, Gresham College
MATTHEW LEITCH | Z/Yen Group
DIONYSIOS DEMETIS | Lecturer in Management Systems, Hull University Business School

ABSTRACT

Cryptocurrencies have the potential to become effective currencies that give a higher level of macroeconomic control, thanks to the information that is available about holdings and transactions, and the potential for automated control mechanisms. However, these cryptocurrencies need to be designed properly and tested before launch. This paper reports the early results of an economic model that simulates a variety of behaviors by economic agents and some simple control mechanisms. An economic simulation model is likely to be a valuable tool in developing effective cryptocurrency systems and interacting with regulators.

[1] DasCoin, along with Z/Yen Group, supported this research financially. Michael Mathias and Terry O'Hearn of DasCoin provided comments on early drafts of the research. None of the authors have an interest in DasCoin or hold its cryptocurrency.

1. INTRODUCTION

Several hundred cryptocurrencies have been launched [Hileman and Rauchs (2017)], with others in the pipeline. Only a few have been successful enough to become widely known and easily exchanged for fiat currencies. Some of these currencies have had value for a while, then lost it. For reasons discussed below, even Bitcoin, the most famous of all, still does not fulfill all the traditional economic functions of money.

Nevertheless, cryptocurrencies have the potential to function as currencies, to revolutionize payments, and to transform finance for the better. From a macroeconomic point of view, cryptocurrencies offer the possibility of currencies whose exact supply is known at all times, along with the exact distribution of holdings of the currency and even the distribution of transaction values. It should be possible to exploit this information to manage the currency more effectively, mostly through automatic control mechanisms that operate quickly and free from political influence.

To achieve this, cryptocurrencies need more than just enthusiastic promotion. They need to be properly designed to function effectively as currencies. However, a cryptocurrency and its users form a complex system and design of its control mechanisms is difficult. To tackle this difficulty, simulation systems could be used. Many designers of complex systems (e.g., environmental, manufacturing, financial, economic) find simulation valuable. The value can derive from improved designs, anticipating problems, or rehearsing reactions to problems. Further, the ability of models to provide a predictive reference (at least short term) makes them useful within a control mechanism [Mainelli (2009)].

Increasingly, adoption of a cryptocurrency may depend on regulatory approval. Another motivation for the economic simulation model presented in this paper was to see if a cryptocurrency simulation might be of use in explaining to a wider audience, including regulators, how a cryptocurrency might perform. The model explored the design issues for cryptocurrencies and the value of using a simulation to test specific design features.

2. DEFINING SUCCESS FOR CRYPTOCURRENCIES

What should a cryptocurrency do well? The objectives could be defined in a number of different ways, but would surely include ideas like popularity, security, availability, efficiency, and speed. On top of those, the objective we focus on here is to be an effective currency.

A cryptocurrency should fulfill the traditional functions of money [Jevrons (1875)]. Exactly what those are has been a topic for scholarly debate for a long time. According to Mainelli (2015), “Money is a technology that communities use to trade debts across space and time.” An old couplet breaks this down in more detail: “Money is a matter of functions four, a medium, a measure, a standard and a store.” This has been analyzed by many authors, including Jevrons (1875).

Some very practical considerations underlie this theorizing. Money should be an effective medium of exchange, enabling two people to make a deal even though they do not have goods or services of equal value to exchange in a barter transaction. The money makes up the difference between what each person offers. This requires that it be accepted over a broad area and time. It should provide a reliable store of value so that if two people make two exchanges separated by a period of time neither feels cheated by the fact that money at one time is worth much more or less than it was at another. To be a standard means holding relative value to a basket of needs, but this is not easy as those needs change with technology, fashion, or scarcity. It should also be supported by a large community of people whose familiarity with the currency and what it can be exchanged for means that they can use it for mentally valuing objects and making decisions, even when the currency is not actually used in a transaction.

A crucial requirement, if a currency is to be effective, is that its value does not change, much, over time. A currency whose value changes greatly day by day cannot reliably store value over time. Losers will feel cheated by value movements. Buyers cannot learn the usual prices for goods they often buy, or shop for good bargains. Sellers cannot advertise prices for goods or services without constantly revising them. Nobody can use the currency in calculations as a proxy for utility.

A stable value in turn requires, among other things, that the money supply should match demand for money.

These requirements are all-embracing. It is difficult to think of any form of money that has achieved all four functions for a significant period of time. To be a store means holding firm over long periods of time, which by inspection has not been attained by fiat currencies nor gold. Fiat currency is a good medium of exchange within a tax zone, but has traditionally been a variable unit of account that leaks value with inflation.

However, these changes in value are slow compared to the rapid fluctuations typical of cryptocurrencies, including Bitcoin. This problem has been presented as a virtue, with cryptocurrencies offered as opportunities for speculation [Bouoiyour et al. (2014)]. Participants are encouraged to buy and hold for a while by prospects of appreciation. It is a game where the winners and losers balance out, except that operators of an exchange or mint take their cut. In theory, at least, a cryptocurrency could be a huge success as a speculative arena despite never being used to buy goods or services. In this role it is similar to online poker, not a currency.

3. AN ECONOMIC SIMULATION MODEL

3.1 Overview

To test ideas for economic control mechanisms for cryptocurrencies before committing them to a live cryptocurrency, it makes sense to build some kind of simulation model. The model described below was designed to focus on economic control, especially money supply and exchange rate mechanisms, not on other potentially relevant considerations, such as energy efficiency, community building, or commercial viability for participants.

It assumes that the cryptocurrency is designed and promoted as a payment system and currency, not as an opportunity for speculation, and that the exchange rate changes will be slow. This is very different to existing cryptocurrencies but realistic for a viable future cryptocurrency. The assumptions underlying the model's design are as follows:

- **A dominant fiat currency:** as the model is to simulate the early stages of a cryptocurrency, the assumption is that there is a dominant fiat currency that is widely used and whose prices are known to users. Everyone uses the fiat currency but they opt in and out of the cryptocurrency. For simplicity, the model assumes only one fiat currency and no inflation, making it a stable reference point for the cryptocurrency.
- **Relatively small scale:** even payments by Bitcoin are on a small scale compared to more established means of electronic payment, so the model assumes that IT costs are not a major factor and that there is no problem operating at the scale arising in the simulations [Croman et al. (2016)].
- **Payments and speculation:** cryptocurrencies are used to pay for goods and services, but users also buy and hold them, hoping to profit from exchange rate changes [Bouoiyour and Selmi (2015)]. This is probably common even among users who are not experienced, skilled currency traders. The model reflects both uses, but the sophistication of speculative trading strategies is very limited.
- **Easy choices between means of payment:** while the purchasing habits of currency users may be stable over time, driven by their basic needs (e.g., food, housing, transport), the choice between alternative payment methods is less constrained. Most users will have an array of alternative means of payment and can choose between them easily the moment before they pay. In the model, users choose between paying with the cryptocurrency and paying with a fiat currency on every transaction and their behavior is sensitive to price differences.
- **Two prices for the same good:** in a society where a cryptocurrency is used alongside a more established fiat currency, goods and services may be offered with prices in each. Sometimes, one of those prices will be a better bargain. For example, if a product is initially given two prices that are equivalent (according to the mid-point exchange rate at that moment), and those prices are not revised for a period of time, then one may become more attractive than the other as the exchange rate changes. This provides a very clear reason for users to pay with either fiat or cryptocurrency on any particular occasion.
- **Unpredictable velocity:** the velocity of a currency is defined as the number of times, on average, that each unit of the currency is used in transactions to buy goods and services in a period of time. The velocity of cryptocurrencies is likely to be especially inconsistent over time because of the easy choice between means of payment and because electronic transfer of funds can be done very quickly. The velocity could be even higher if robots initiate transactions. The model tracks changes in velocity, calculated in both the conventional and in a distribution-adjusted way.

- **Consumer oriented currency exchange:** most people experience currency exchange when going on holiday or on a business trip. They buy the foreign currency they need at the advertised price rather than by putting more complex forms of order into an order-driven exchange. The model uses the simpler form of currency exchange between the cryptocurrency and a fiat currency.
- **Controlled money supply at a price:** the quantity of cryptocoins in issue at any time is precisely known and can be precisely controlled. This is different from the situation with fiat currencies today [McLeay et al. (2014)]. In the model, users can buy newly minted cryptocoins for a price, which approximately represents the situation with at least some cryptocurrencies at the time of writing. In the case of DasCoin, the cryptocoins can be acquired in exchange for Cycles, which themselves are bought with fiat currency or Bitcoin. Indirectly, this establishes an approximate cost of acquiring newly minted cryptocoins. The situation with Bitcoin is different. Bitcoin can be mined and so acquired at a price that reflects the investment in computing power. However, getting started at mining is a significant investment so Bitcoin users mostly do not mine sporadically when it is a cheaper option than going to an exchange. The model begins simulations with an initial stock of cryptocoins held by the exchange, and then tracks the quantity of cryptocoins in issue. The cost to acquire newly minted coins can be varied.

Clearly, a number of features of current and future cryptocurrencies and exchanges that are sometimes important are missing from this model. Some of these are mentioned below as opportunities for future development.

3.2 The simulation cycle

The economic model is stochastic and based on intelligent agents that interact over a sequence of discrete days, buying and selling goods, and adjusting their cryptocurrency holdings by exchanging for the fiat currency and buying newly minted cryptocoins. There have been several examples of agent-based models of currencies [Chatagny and Chopard (2000), Cocco and Marchesi (2016), Delage et al. (2010), Setzu (2007), Usami et al. (2006)]. The model is implemented as an R script. A wide range of parameters can be set to control the behavior of the model.

The agents in the model are: (1) merchants, who offer goods for sale; (2) customers, who buy those goods; and (3) an exchange market maker that buys and sells the cryptocurrency and fiat currency.

Each day follows the same pattern, as follows: (1) merchants decide if they will use the cryptocurrency and, if so, how often they will revise their prices; (2) merchants adjust some cryptocurrency prices; (3) customers decide if they will use the cryptocurrency; (4) customers make purchases of goods from merchants and, each time, decide if they will pay with cryptocurrency or fiat currency; (5) customers and merchants decide how much cryptocurrency they wish to hold and adjust their holding by buying from or selling to the currency exchange, or by buying newly minted coins; and (6) the currency exchange market maker decides what prices for the cryptocurrency to set for the next day.

> "To test ideas for economic control mechanisms for cryptocurrencies before committing them to a live cryptocurrency, it makes sense to build some kind of simulation model."

The fiat currency prices of goods remain fixed throughout each trial, but merchants can set cryptocurrency prices too. Merchants do this by using the mid-point exchange rate for the day to set a price that is equivalent to the given fiat currency price. They can either revise their prices daily, weekly, or every 30 days. The decision to start or stop using the cryptocurrency is randomized, as is the choice of frequency for revising prices. However, the decision to use the cryptocurrency is influenced by the apparent success of the currency and the amount of positive publicity around it. Once merchants have started to use the cryptocurrency they are encouraged to continue by their sales in the cryptocurrency.

Customers decide to use or not use the cryptocurrency in a similar, randomized way. Once they start using it they are encouraged to stay by the savings they make through cryptocurrency purchases.

Customers do not shop around for alternative suppliers of the same goods or services. However, customers who are cryptocurrency users will consider any cryptocurrency prices offered for goods they want to buy and will decide how to pay. This is based on choosing the cheapest way to pay given the two advertised prices, the day's midpoint exchange rate, and the cost of buying newly minted cryptocoins. (The midpoint exchange rate is the geometric mean of the bid and ask prices.)

Merchants and customers decide the amount of cryptocurrency they would like to hold at the end of each day by using the same basic strategy, but with parameters that are randomized between agents so that some heterogeneous behavior results. The users are assumed to have a cash amount (specified in fiat currency) that they hold at all times and allocate between fiat currency and cryptocurrency. Following the Kelly Strategy [Kelly (1956)], they allocate this cash amount according to their probability of each being the better investment. For example, if the user thinks that the cryptocurrency is 60% likely to appreciate relative to the fiat currency then the user will decide to hold 60% of the cash amount in cryptocurrency.

The perceived probability of cryptocurrency being the best investment is driven by several variables and in all cases the user considers the recent trend of changes in those variables [Izumi (2010)]. For example, if overall holdings of the cryptocurrency have been rising on most days recently then the user will take that as an encouraging sign and want to hold more.

The perception of commercial activity and exchange activity is tempered by knowledge of the distribution of that activity. This is modeled by having the agents react to market indicators multiplied by the relative entropy[2] of the distribution of the transactions or holdings involved. For example, if a lot of cryptocurrency is held, but only by one person, then this is little better than no cryptocurrency being held at all, while the same quantity of cryptocurrency equally distributed across many users is a much more encouraging sign of a viable community of users.

The model has several alternative strategies for revising the exchange rate of the cryptocurrency for each day. These are discussed in more detail below, where their effect is illustrated. However, the model does not fully reflect possible shortages of demand or supply that might mean the exchange cannot meet all orders.

Figure 1: Exchange rate (FC/CC) over time in a typical simulation trial (the rate is capped at 1.2 by the cost of minting).

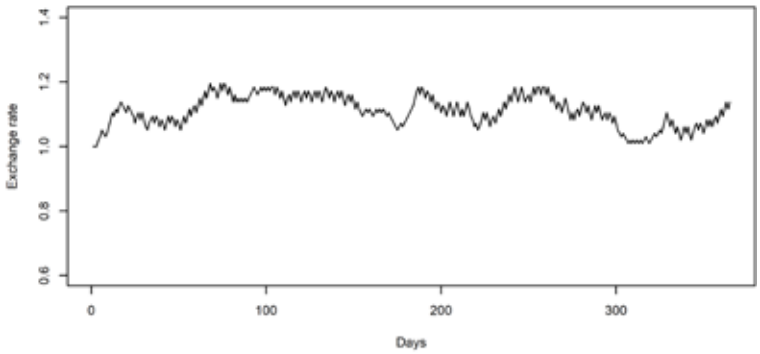


Figure 2: Minting is sporadic and occurs when the cost of buying newly minted cryptocoins is less than the cost of buying cryptocoins on the exchange.

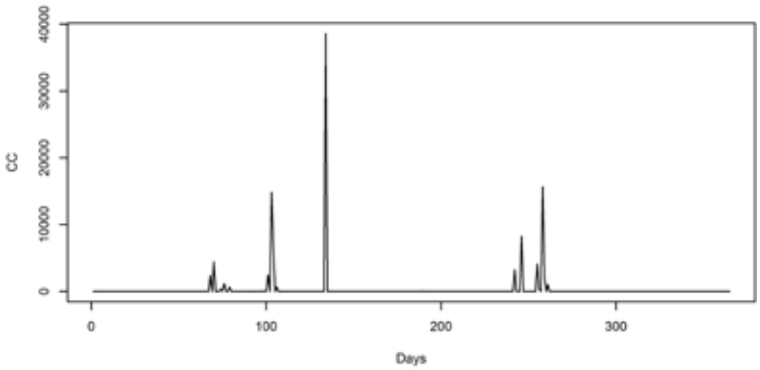
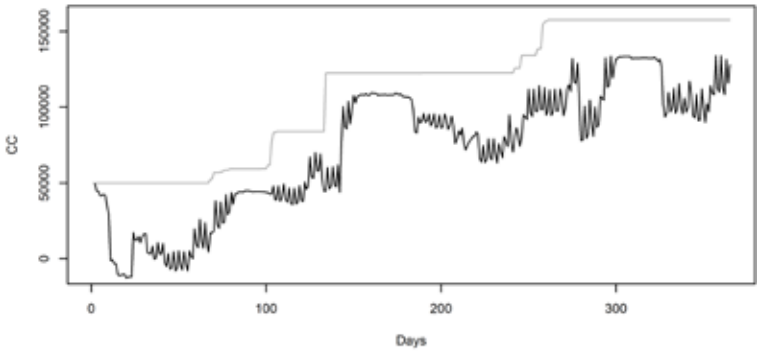


Figure 3: When cryptocoins are minted, the total quantity of cryptocoins in existence increases (the grey line) and this tends to allow the pool of cryptocoins held by the exchange market maker (black line) to increase.



[2] Relative entropy was defined as the entropy of the distribution divided by the entropy of a uniform distribution of the same total value. For a sequence of N non-negative values bi, i = 1..N, Relative Entropy = $-\frac{\sum_{i=1}^{N}\frac{b_i}{N}\times\log_2(\frac{b_i}{N})}{\log_2(N)}$, with 0 x $\log_2(0) = (0)$

3.3 A typical trial

As discussed later in this article, many aspects of the behavior of a new cryptocurrency are unpredictable and sensitive to details of agents' decisions. However, some features of trials with the model are fairly consistent and are visible in the following, typical example. This provides some context for understanding the variations and effects of control mechanisms discussed later.

This was a trial simulating 365 days with 10 merchants, 40 customers, and a cost of minting that was 1.2 times the initial exchange rate of the cryptocurrency (CC), which was 1 unit of the fiat currency (FC). The exchange rate evolved as shown in Figure 1, clearly capped by the minting cost of 1.2. Minting occurs when the rate to buy CC rises and hits the minting cost, as shown in Figure 2.

Although minting is only sporadic, it has a great effect on the total quantity of CC that exists, relieves pressure on the exchange's pool of cryptocurrency, which was becoming depleted, and constrains the exchange rate.

With the minting cost set much higher the exchange rate tends to rise much further before being capped, but exactly what happens depends on other features of the system. If new cryptocoins were to be mined continuously in a way that was largely unrelated to demand rather than purchased from the mint then the effects would depend on many features of the system but would be less controlled.

The demand for CC is largely driven by the gradual rise in users, as seen in Figure 4.

The holdings of users (excluding the exchange market maker) also rise, but not as smoothly. Figure 5 shows these holdings, but multiplied by the relative entropy of the holdings. Relative entropy is a number between 0 and 1 that reflects inequality in the distribution. A relative entropy of 1 occurs when all holdings are of equal value. A relative entropy of 0 occurs when only one user holds CC. (The model also tracks this quantity using the Gini Index[3] as a measure of inequality.)

Use of CC to buy real goods also increases, but is sporadic, as shown in Figure 6. Comparison with the exchange rate time series reveals that the activity corresponds to periods of falling or stable prices.

Figure 4: The total number of users (both merchants and customers) rises rapidly at first, then slows down and may decrease.

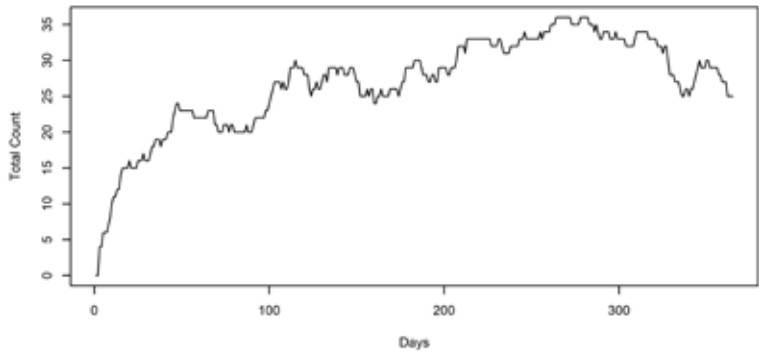


Figure 5: Holdings of CC (excluding the exchange market maker) rise, but not smoothly. The plot shows the total holdings of CC by merchants and customers multiplied by the relative entropy of those holdings so that both the quantity and distribution of CC holdings is considered.

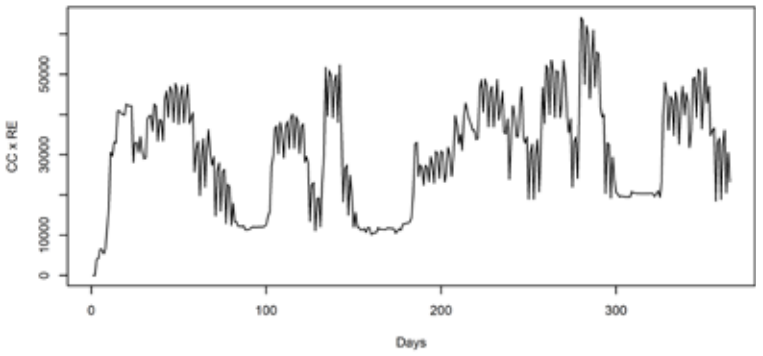
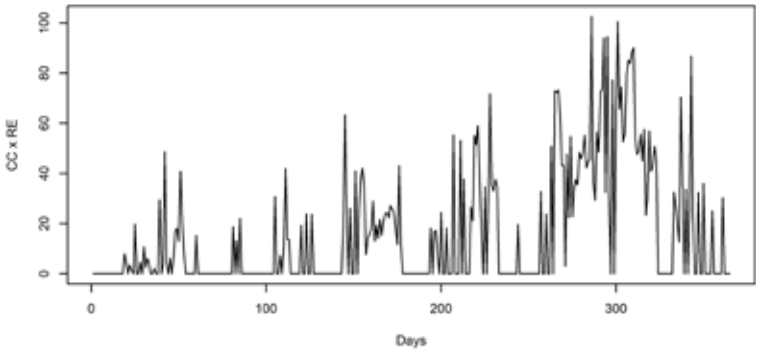


Figure 6: Use of CC to buy real goods and services increases but is sporadic.



[3] The Gini Index is 1 for complete inequality and 0 for complete equality, which is the opposite of Relative Entropy. The model tracks and uses 1 - G, where G is the index.

Commercial activity is driven by a falling rate, as shown by a scatter plot of the moving average of daily exchange rate movements against commercial activity. In Figure 7, it is clear that commercial activity appears most strongly when the trend is flat or negative.

If the model was more sophisticated and customers shopped around for the best deal on identical products from alternative suppliers then the effect of these flurries of CC spending would be stronger, with merchants offering CC prices making more sales.

3.4 Different behavior from varying assumptions about agents' decisions

Future cryptocurrencies may attract, and prompt, different behaviors among users. They might vary in being market followers or contrarians, in having a long- or short-term perspective, or in being more or less susceptible to hype. This probably means that it is not feasible to predict, accurately, the evolution of a particular cryptocurrency. However, it should be possible to simulate a variety of plausible behaviors and study how the control mechanisms perform in the face of challenging patterns.

To accommodate this, the simulation model has a large number of parameters that affect its behavior. Many of these concern the decisions of agents. Two important examples are assumptions about decisions to get involved with the cryptocurrency and decisions on how much of it to hold each day.

A very common feature of simulation trials with the model is a rapid initial uptake of the CC as merchants and customers decide to use it and begin to hold stocks of the CC. This demand alone drives the price up.

However, after an initial rise the pressure is reduced as the number of customers and merchants opting in reduces to match the increasing number opting out. Although the total population of merchants and customers in the model is intended to represent only those people who would ever be interested in using a cryptocurrency, this is still not realistic. In the real world, there are billions of potential users but only a relatively tiny proportion of them become actual users. A cryptocurrency could perhaps rise as a result of recruiting new users for a long period of time.

Figure 7: The trend of exchange rate changes (represented by the exponentially weighted moving average of daily differences) is linked to purchasing real goods and services. It is most common when the currency falls, making prices quoted in CC more attractive than FC prices, until prices are adjusted.

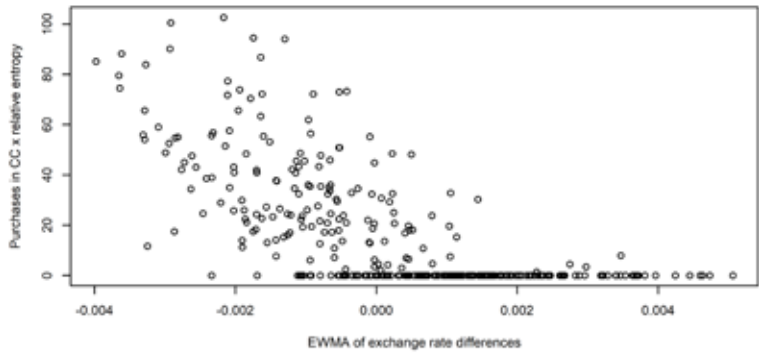


Figure 8: Increasing the population of potential customers from 40 to 400 and reducing the propensity to become users gives a smoother growth of user numbers, reaching a plateau determined by assumptions about joining and leaving.

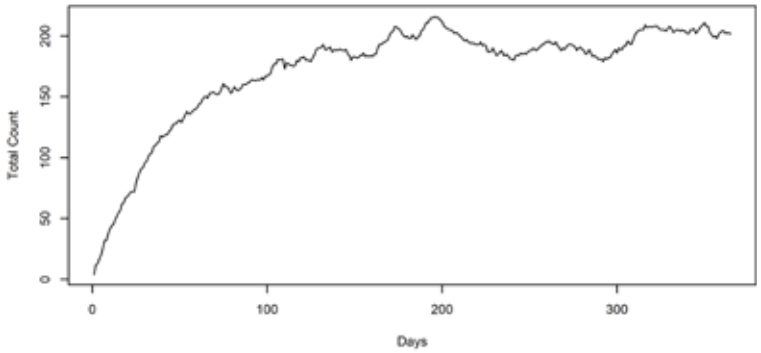
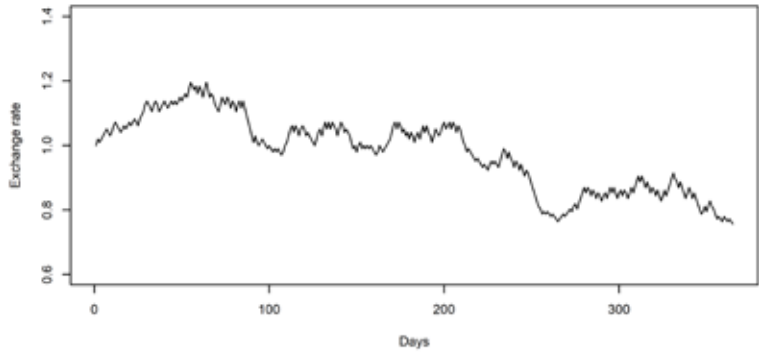


Figure 9: With relatively unreactive, users the exchange rate tends to hover in between sudden jumps.



Increasing the pool of potential users, but decreasing the proportion of them that join each day, produces simulations that are more predictable in percentage terms because of the larger numbers involved (e.g., Figure 8). However, the total fraction of the population that become users depends entirely on the assumptions made about how people will respond to promotional activity, news of the progress of the cryptocurrency, and so on.

In these early model simulations, the time taken to reach the approximate equilibrium level where joiners equal leavers does not change much as the total population is increased. This is a surprise, but perhaps reflects the assumption that the entire population is exposed to information about the cryptocurrency at the same time. In reality, perhaps people pay attention to this news only occasionally and there is some kind of spreading awareness that slows the process. Alternatively, it may be that salespeople promoting cryptocurrencies and services related to them take time to work through the population of potential buyers.

Another example of sensitivity to assumptions is the effect of making users more reactive to recent information. In the model, the extent to which users react to the latest information rather than wait to see if trends persist is controlled by the individual recency factors of each user and each variable. However, by constraining those into narrow ranges it is possible to see the effect of making everyone generally more or less reactive.

Figures 9 and 10 contrast the typical appearance of the exchange rate time series with low and high reactiveness, respectively. With low reactiveness, the series has periods of small rises and falls, with rapid changes of direction, interrupted by occasional big rises or falls. With high reactiveness, the series is more often characterized by a more even see-saw rise and fall with few dramatic changes. This change is not reflected much in the change in standard deviation of daily differences, which goes from 0.1534 to 0.1555, nor in the Fractal Dimension[4] of the time series, which rises from 1.676 to 1.699. However, it is clear that the distribution of runs up and down has changed, with many more movements of around 0.05 in size.

Figure 10: With relatively reactive users the exchange rate tends to rise and fall more continuously, giving the plot a serrated appearance.

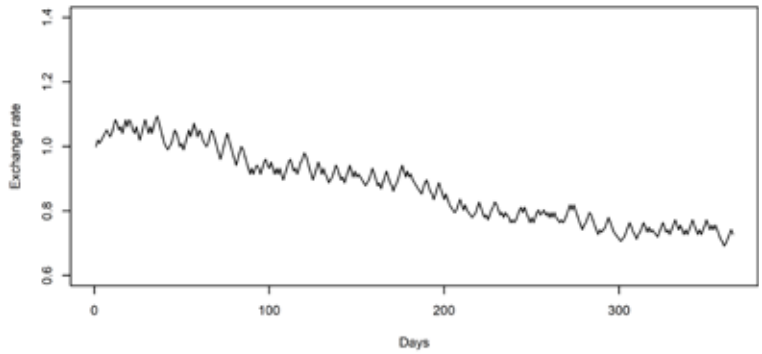


Figure 11: With fairly even distribution of goods, purchasing, and money, the time series of CC holdings multiplied by relative entropy has a complex shape with many rises and falls.

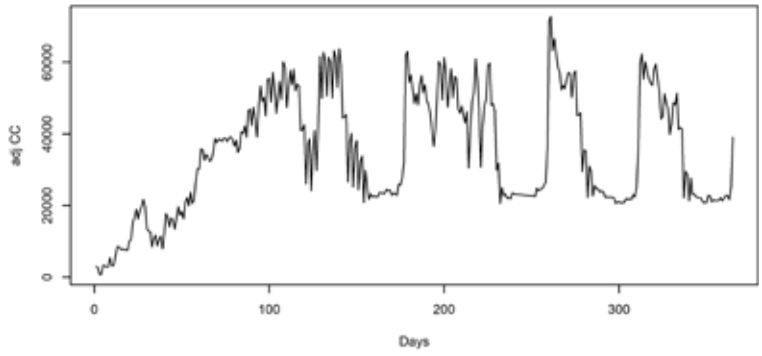
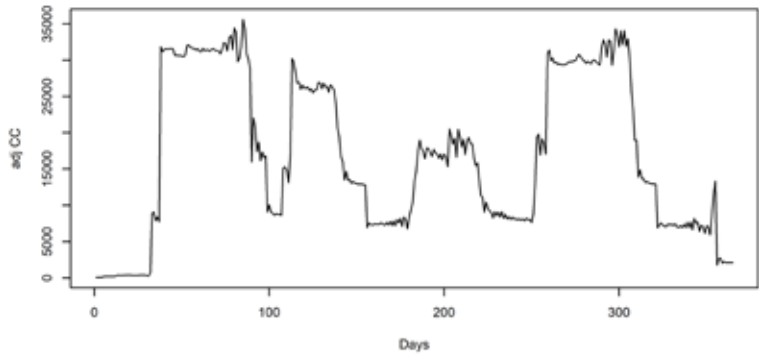


Figure 12: With unequally distributed goods, purchasing, and money, the time series of CC holdings multiplied by relative entropy has a simpler shape, dominated by large rises and falls.



[4] The Fractal Dimension values were calculated using a refinement of Higuchi's algorithm [Cervantes-De la Torre et al. (2013)].

The distribution of holdings and activity is also important. It is possible to vary this while keeping the total number of merchants, customers, and goods constant, and keeping the average number of purchases per customer per day constant along with the average price of goods. With the goods and money fairly evenly distributed, the relative entropy of holdings of CC emerges at around 0.9 and the graph of CC holdings multiplied by relative entropy typically shows a fairly complex shape with many periods of rapid alternating rises and falls, mixed with some large jumps (Figure 11).

In contrast, with very unequal distributions of goods, purchasing, and money, the relative entropy emerges at around 0.2 and the plot of CC holdings multiplied by relative entropy is simpler, with sudden jumps but less other activity (Figure 12).

3.5 Control mechanisms

The rule used by the exchange market maker to update the exchange rate for each next day is a highly influential control mechanism. It is not true to say that any rule that adjusts the rate up a bit when demand exceeds supply, and adjusts it down when supply exceeds demand, will have roughly the same effect thanks to a natural negative feedback loop.

With the same starting conditions but changing only the rule for updating the exchange rate, very different results are obtained. A further series of plots illustrates the effect on exchange rate using a simulation in which the cost of minting is set very high so that the exchange rate is not capped and the money supply is fixed. With a minimal rule that adjusts the rate up by 1% or down by the same multiple the result is shown in Figure 13.

With a rule that adjusts the rate more when the absolute value of the net demand for CC is larger, a visibly different time series results (Figure 14). Bursts of demand for CC are met by a rapid increase in the rate that then subsides, nearly as rapidly.

The same rule but with a weaker reaction to differences between supply and demand produces a more stable exchange rate (Figure 15) with a narrower range but similar characteristics. Note the slight upward trend.

Finally, another less reactive rule, but this time with a tendency to avoid the market maker's pool becoming depleted or excessive (Figure 16). The exchange rate now keeps returning to the original value of 1, even though this is not an explicit part of the rule used.

Figure 13: The exchange rate over time with a minimal rule for adjusting the rate each day.

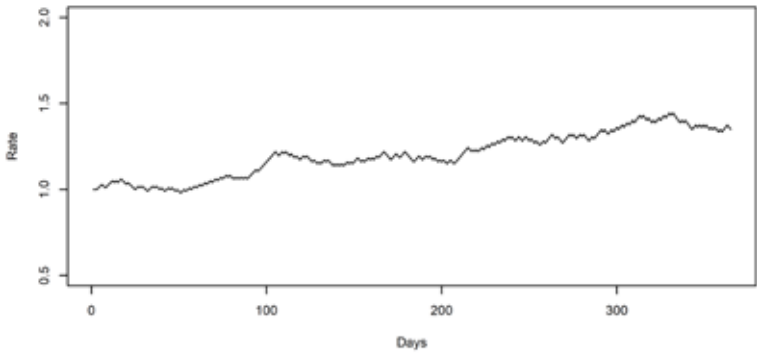


Figure 14: With a rule that adjusts the rate more as the difference between demand and supply for the CC increases, a greater overall variation in rates is produced, and with a different quality of variation.

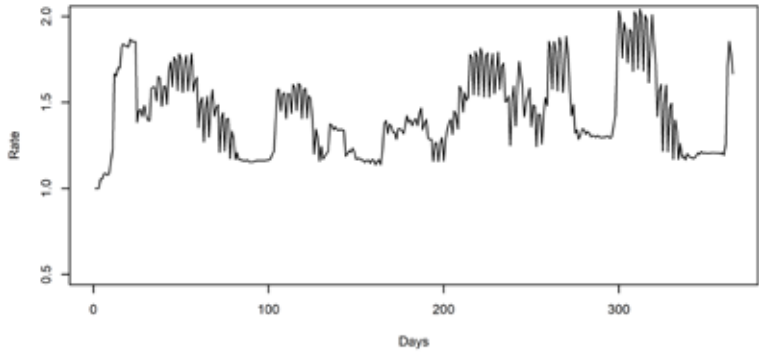
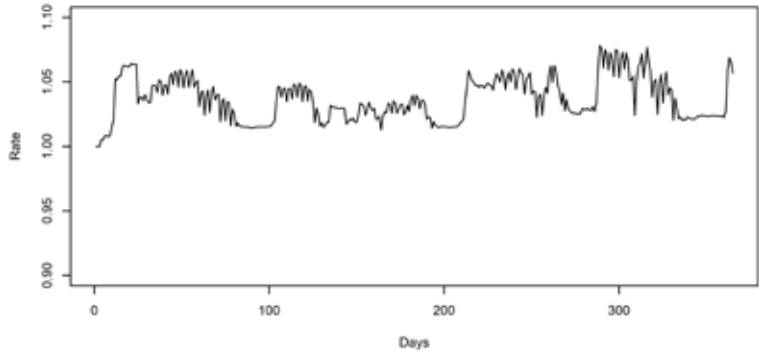


Figure 15: With a rule that reacts less strongly to an imbalance between demand and supply the variation in rates is much less, though the patterns of variation appear similar. Note the narrower scale on the vertical axis compared to Figure 14.



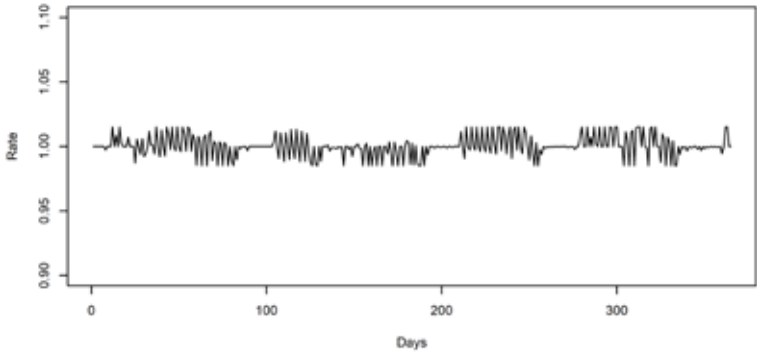
4. AREAS FOR FUTURE RESEARCH

The sensitivity of the model to different assumptions about agents' decision making and other factors that are unlikely to be predictable before a cryptocurrency is launched strongly suggests that exact prediction of cryptocurrency behavior is unlikely to be feasible before going live. However, it might be valuable later and could even be a part of a control mechanism. However, by simulating a variety of plausible behaviors it is possible to test adaptive management strategies for managing cryptocurrencies and demonstrate the information that could be available to regulators and governments. The model needs to be able to simulate a variety of behaviors, including potentially destructive loops and catastrophic changes, and allow alternative control rules to be tested.

A number of potentially interesting effects and features of cryptocurrencies and their environments could be incorporated into future developments of the simulation model. These include the following:

- Transaction costs
- International use, where the cryptocurrency might be an alternative to two or more fiat currencies
- Very rapid transactions within a single day, perhaps also driven by algorithmic trading, that might lead to large movements in exchange rates within a single day
- Any increased tendency to set and advertise prices in the cryptocurrency when the exchange rate is stable
- Other reasons for using the cryptocurrency, such as social display, to feel up-to-date, or to facilitate crime
- The influence of social networks in deciding who gets enthused about a cryptocurrency and when
- The impact of focused, energetic sales effort that persuades particular subgroups of the population to participate rather than just pushing sales information at everyone
- Highly damaging news stories, such as stories of hacking and arrests, that might deter people from using the cryptocurrency
- Complex, idiosyncratic features of cryptocurrency designs, whose effect is often to complicate decision making for users and increase the uncertainty involved for all participants.

Figure 16: With a rule that also tries to keep the market maker's pool within a range the exchange rate is also constrained and repeatedly returns to the original value of 1.



The economic usefulness of a cryptocurrency depends to a large extent on how evenly distributed it is. If a handful of people own nearly all of it, then, even if there are many people with non-zero holdings, its usefulness will be limited.

A particular feature of CCs going forward may be their ability to test empirically the Quantity Theory of Money, taking distribution into account. The distinction between "distribution of activity" and "distribution of holdings" suggests an extension to the Quantity Theory of Money, where MV = PT, the "Fisher Equation" [Fisher (1922)], where M = money supply; V = velocity of circulation (the number of times money changes hands), P = average price level, and T = volume of transactions of goods and services.

The extension might be along the lines of d(H)MV = d(A)PT where d is a distribution or entropy measure for "holdings" and "activity."

Management strategies need not be restricted to minting and exchange rules. Action could also be taken to link the currency to real goods and services by, for example, offering a catalogue of products with stable cryptocurrency prices.

Other important issues for further research include:

(1) Comparison with alternative electronic payment methods, many of which are highly efficient and more secure than credit and debit cards.

(2) The economic arguments for and against currencies with a small user base, such as the local currencies of Germany discussed by Rösl (2006) and by Z/Yen (2011).

5. CONCLUSION

Cryptocurrencies in the future have the potential to contribute efficiency and economic control, but better designs are needed and these will need to be tested. Simulation is a good way to do this before committing to a live cryptocurrency and might also help with control once the cryptocurrency is live.

Exploring such a model's behavior has confirmed that results are sensitive to the detailed characteristics of agents' decisions, and are unpredictable. This is seen in the response to changing the number of customers, the reactiveness of users to routine news of the cryptocurrency, and the way goods, purchases, and money are distributed across users.

Powerful controlling effects can be achieved by adjusting the cost of newly minted cryptocoins and by adjusting the exchange-rate price revision rule. Almost certainly, other sources of unpredictability and of control can be found. This is just the start of an exciting line of research.

All this suggests that designers of cryptocurrencies should develop and test their design (through simulation and mathematical analysis), including any exchange facilities, and should focus on rules that adapt to events rather than being fixed, based on initial assumptions. It may never be possible to predict in advance the evolution of a cryptocurrency, but it should be possible to develop a model that can be used to test control mechanisms against a wide range of factors and effects.

References

Bouoiyour, J. and R. Selmi, 2015, "What does Bitcoin look like?" Annals of Economics and Finance 16:2, 449-492

Bouoiyour, J., R. Selmi, and A. Tiwari, 2014, "Is Bitcoin business income or speculative bubble? Unconditional vs. conditional frequency domain analysis," MPRA, http://bit.ly/2GC0SVp

Cervantes-De la Torre, F., J. I. González-Trejo, C. A. Real-Ramírez, and L. F. Hoyos-Reyes, 2013, "Fractal dimension algorithms and their application to time series associated with natural phenomena," Journal of Physics: Conference Series 475:1

Chatagny, R. and B. Chopard, 2000, "A parallel model for the foreign exchange market," Parallel Computing 26:5, 587-600

Cocco, L. and M. Marchesi, 2016, "Modeling and simulation of the economics of mining in the Bitcoin market," PLoS ONE 11:10, http://bit.ly/2ogTHuY

Croman, K., C. Decker, I. Eyal, A. E. Gencer, A. Juels, A. Kosba, A. Miller, P. Saxena, E. Shi, E. G. Sirer, D. Song, and R. Wattenhofer, 2016, "On scaling decentralized blockchains," in Clark J., S. Meiklejohn, P. Ryan, D. Wallach, M. Brenner, and K. Rohloff, (eds.), Financial cryptography and data security. Lecture notes in computer science, vol. 9604, Springer

Delage, V., C. Brandlhuber, K. Tuyls, and G. Weiss, 2010, "Multi-Agent based simulation of FOREX exchange market," Maastricht University, Department of Knowledge Engineering, the Netherlands Teramark Technologies GmbH, Munich, Germany

Fisher, I., 1922, The purchasing power of money, its determination and relation to credit, interest and crises, assisted by Harry G. Brown, New and Revised Edition, Macmillan, original publication 1911

Hileman, G. and M. Rauchs, 2017, "Global cryptocurrency benchmarking study," working paper, doi:10.2139/ssrn.2965436

Izumi, K., 2010, An artificial market model of a foreign exchange market, Economic Web Institute, http://bit.ly/2HDrNRJ

Jevrons, W. S., 1875, Money and the mechanism of exchange, D. Appleton and Co.

Kelly, J. L., 1956, "A new interpretation of information rate," Bell System Technical Journal 35:4, 917-926

Mainelli, M., 2009, "Environmental consistency confidence: scientific method in financial risk management," in Tarantino, A. and D. Cernauskas (eds.), Risk management in finance: six sigma and other next-generation techniques, John Wiley & Sons, 273-288

Mainelli, M. R., 2015, Unblock the shared economy, Duke Corporate Education

McLeay, M., A. Radia, and R. Thomas, 2014, "Money creation in the modern economy," Bank of England Quarterly Bulletin 2014 Q1

Rösl, G., 2006, "Regional currencies in Germany – local competition for the Euro? No. 43." Deutsche Bundesbank Discussion Paper, Series 1: Economic Studies, Frankfurt

Setzu, A., 2007, "A framework for financial markets modeling and simulation," Exchange Organizational Behavior Teaching Journal

Usami, A., R. Tsuya, T. Iba, and H. Takayasu, 2006, "Building a simulation model of foreign exchange market," The 5th International Conference on Computational Intelligence in Economics and Finance (CIEF2006), Taiwan

Z/Yen, 2011, "Capacity trade and credit: emerging architectures for commerce and money," report prepared for the City of London Corporation, ESRC, and Recipco

Narrow banks and fiat-backed digital coins

ALEXANDER LIPTON | Connection Science Fellow, Massachusetts Institute of Technology (MIT), and CEO, Stronghold Labs
ALEX P. PENTLAND | Toshiba Professor of Media Arts and Sciences, MIT
THOMAS HARDJONO | Technical Director, MIT Trust::Data Consortium, MIT

ABSTRACT

We outline a framework for issuing fiat-backed coins to a wide set of end-users. We show that a narrow bank is an important part of this framework, needed to increase circulation and acceptance of such coins. We argue that fiat-backed coins issued by a purpose-built narrow bank have considerable advantages compared to central bank digital cash, and can be used to achieve improved financial stability and solve some of the more vexing problems affecting financial infrastructure.

1. INTRODUCTION

This article describes the concept of a fiat-backed digital coin (FBDC) and marries it with the idea of a narrow bank (NB). It outlines an approach to increasing FBDC acceptability and circulation from a small set of initial sponsors to a much wider (but still limited) group of potential users, such as small- and medium-sized enterprises (SME) and individuals, via a purpose-built NB. In short, the idea is to apply distributed ledger technology (DLT) to give a new lease of life to the old NB concept, and to use a NB as a centerpiece (glue) at the heart of a digital ecosystem. When properly designed, a NB can be used for several related purposes including issuance of FBDC. While we describe the concept of a NB in detail below, it is worth mentioning that such a bank has (almost) perfectly matching assets and liabilities, so that it is impervious to market and liquidity risks. In a nutshell, on its asset side, a NB has only central bank cash, or short-term government obligations, while on its liability side it has deposits and equity. In the old days, the assets would be solely in gold, later on a combination of gold and paper money, and finally, in our time, the assets would predominantly be electronic balances on deposit with the central bank.

While the idea of a NB is not new, it is not clear if a truly NB had been ever built. Currently, almost all banks are fractional reserve in nature, and are engaged in maturity transformation by maintaining long-term assets and short-term liabilities, thus opening themselves to risks of potential runs and other hazards, up to and including default.

We share the view succinctly expressed by Aristotle: "But money has been introduced by convention as a kind of substitute for need or demand; and this is why we call it money (νομισμα), because its value is derived, not from nature, but from law (νομοζ), and can be altered or abolished at will" (Aristotle, The Nicomachean Ethics.) In view of this quote, we wish to design FBDC in a manner compliant with all applicable laws, including the know-your-customer (KYC) and anti-money-laundering (AML) requirements.

FBDC, being a digital currency, naturally resides on a purpose-built distributed ledger. By now, building a distributed ledger system, which can function without a central authority, is well understood. Bitcoin, first described by S. Nakamoto (2008) in the seminal white paper, inspired the creation of more than a thousand of other cryptocurrencies, all with various degree of novelty and utility (if any). By construction, these currencies are native tokens, residing on a blockchain, and, as such, can be controlled by the agreed-upon consensus mechanism among agents maintaining and operating such a distributed ledger. However, until now, attempts to make real-world assets, first and foremost, fiat currencies, to be properly incorporated into a blockchain have been unsuccessful.[1] Yet, without a satisfactory solution to this all-important problem, it is not possible to make blockchains a part of the mainstream payment infrastructure.

We argue that for a consortium of sponsors (such as large banks), who are satisfactorily vetted in advance and able to pass the KYC and AML requirements, a fiat currency can be digitized with the assistance of the corresponding central bank, who agrees to convert some of the participating banks' reserves into digital tokens on a one-to-one ratio. This is the approach taken by Clearmatics, a software company based in London.[2] However, for a larger group of potential users, including, in addition to the original consortium member banks, some non-banking financial institutions, as well as SMEs and, possibly, individuals, direct participation of the central bank becomes problematic. We propose a solution, which boils down to building a special-purpose NB, whose operations are streamlined and safeguarded as much as possible in order to limit operational risks. This bank will keep fiat currency submitted by the users and issue digital tokens in return. These tokens will circulate within the group of users in a fast and efficient manner by utilizing distributed ledger mechanism, thus creating native tokens convertible into fiat currency at will. We emphasize that operational risks are always present, but this is true not only for the setup we are proposing, but for ordinary cash and bank deposits too, and, in all probability, to a larger degree.

2. DISTRIBUTED LEDGERS AND CRYPTOCURRENCIES

2.1 Background

For decades, little or no attention was paid to the infrastructure supporting the internal workings of the financial ecosystem. As a result, this infrastructure dramatically fell behind the actual demands of the marketplace. This fact became completely obvious during the global financial crisis (GFC), which put enormous stresses on the transactional infrastructure and pushed it almost to the breaking point. Currently,

[1] Tether is a representative example of such an attempt.
[2] The lead author is a member of their advisory board.

financial infrastructure is centered around private centralized ledgers maintained by individual banks, which are reconciled through the central banks' ledgers [see, e.g., Norman et al. (2011)].

Although for centuries this system served finance reasonably well, it has always been plagued with numerous issues, related to both domestic and foreign transactions. In the current framework, even simple cash transfers (not to mention transactions involving securities) are slow and, under certain circumstances, risky.

In Figure 1 we show a typical domestic bank transaction between Alice and Bob who have accounts at two different banks.

In Figure 2 we show a typical cross-border transaction between Alice and Bob who have accounts at two different banks located in their respective countries.

Fortunately, remarkable technological breakthroughs – mostly related to cryptocurrencies, distributed ledgers, and related concepts – simultaneously focused attention of key decision-makers and technical experts on the glaring need for transforming the financial infrastructure, and, at the same time, indicated how such a transformation can be accomplished.

2. DISTRIBUTED LEDGER DESIGN

2.2.1 Public versus private ledgers

A distributed ledger can be designed along several lines. The key question is whether a distributed ledger is needed in the first place. If the answer is affirmative, then two other questions need to be answered: (A) should the ledger be made permissionless or permissioned, or, equivalently, public or private; and (B) who, and via which mechanism, maintains its integrity. We feel that the FBDC carrying ledger should be semi-permissioned, so that everyone should be able to join, but participants should be known to the NB at the very least when they exchange fiat currency for tokens and, conversely, when they exchange tokens for fiat currency. In the interim, the participants probably can retain anonymity, even though the exact degree of anonymity is open to debate. It is clear that participants' identities have to be anonymous to other users; however, lawful legal authorities, under limited and well-defined conditions, should be allowed to uncover the true identities of participants.

2.2.2 CONSENSUS MECHANISMS

Given that different actors, whose interests are not aligned, are participants of the distributed ecosystem, it is imperative to design a mechanism for achieving consensus among them. Such a mechanism has to be able to tolerate Byzantine faults, both intentional and unintentional, as discussed by Castro and Liskov (1999), Lamport et al. (1982), and many others.

So far, the most successful practically implemented consensus mechanism is based on the competitive proof of work (PoW) [see Nakamoto (2008)]. However, by its very nature, this mechanism consumes enormous amounts of energy and is not suitable for large-scale applications. Accordingly, other options, including proof of stake, proof of burn, proof of age, and random selection of validators, have to be considered [see, e.g., Buterin (2013), Micali (2016)].

Figure 1: A sketch of a transaction between Alice and Bob, in which Alice sends Bob U.S.$100

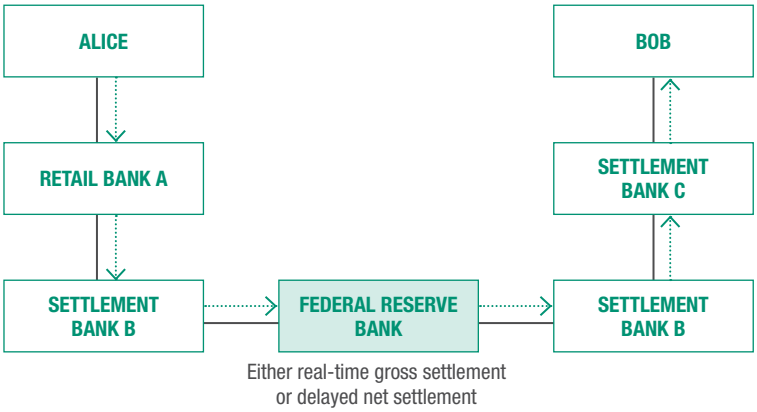


Figure 2: A sketch of a transaction between Alice and Bob, in which Alice sends Bob £100

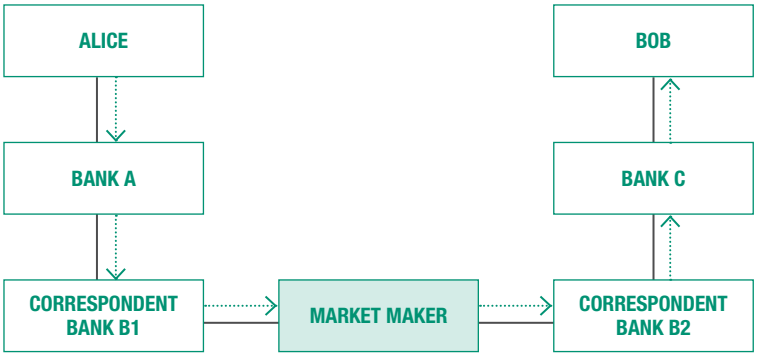
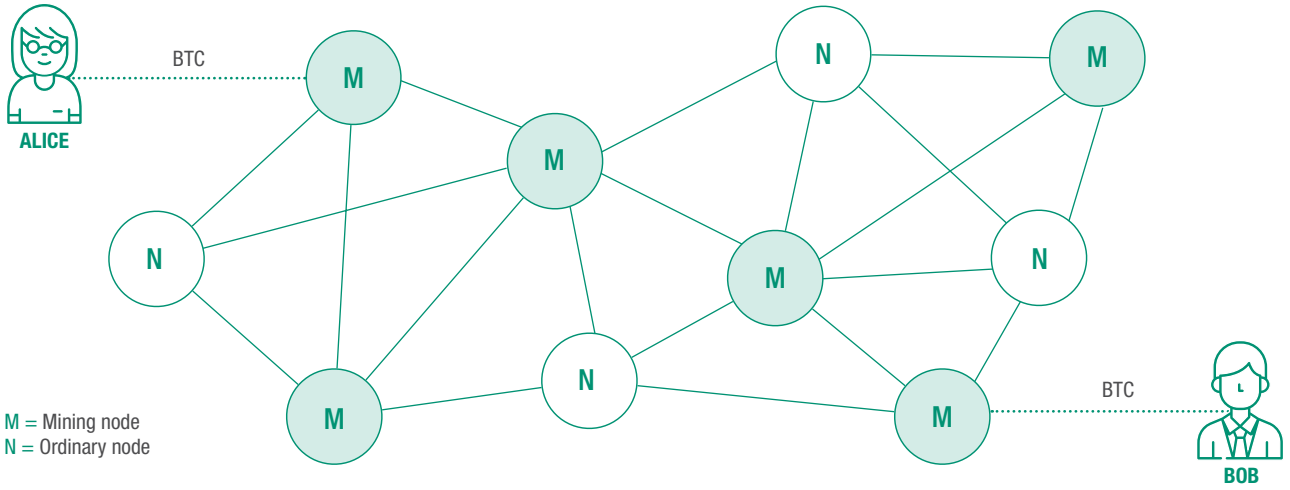


Figure 3: A sketch of a transaction between Alice and Bob, in which Alice sends Bob a BTC



For the large-scale applications, we are leaning toward using validators or notaries, running full nodes and verifying transactions along the lines of majority votes, as, for example, done in the Ripple protocol [Schwartz et al. (2014)].

2.3 Bitcoin setup

Recently, DLT attracted a lot of attention from both the industry and the general public. Astonishing success of Bitcoin demonstrates that a distributed ledger without central authority can function in a coherent and Byzantine fault tolerant fashion in real life. While very impressive from a technical standpoint, in its original form Bitcoin is not suitable for high finance. The reasons are simple – the system is pseudonymous, does not solve the all-important KYC and AML requirements, is not scalable by design as its throughput speed is no more than seven transactions-per-seconds (TpS), and consumes enormous amounts of electricity. Moreover, the volatility of Bitcoin is very high, which precludes it from being useful for transactional purposes, not to mention for lending and borrowing. Some observers even argue that the dominant raison d'être of Bitcoin is to facilitate illegal activities.[3] In addition, by construction, Bitcoin is a native token, which lives on the distributed ledger, while fiat currencies and other financial assets do not reside there. As a result, Bitcoin cannot solve the delivery versus payment (DvP) problem. While in theory it is easy to move Bitcoin from one address, represented by a public key to next, it is not possible at all to ensure the movement of currency, goods, and services in the opposite direction. Since there are no laws regulating these movements, the whole system is prone to all kind of malfeasance. In Figure 3 we show a typical transaction between Alice and Bob who have pseudonymous Bitcoin accounts identified by their public keys.[4]

2.4 FBDC setup

Having said that, Bitcoin setup can be used as a prototype for building a distributed ledger more suitable for interbank transactions and other financial purposes. Several issues need to be resolved before this goal can be achieved:

- The ledger has to be made at least semi-private (if not private) in order to meet KYC requirements.
- A right balance has to be struck between privacy and accountability in order to satisfy the AML requirements.
- An industrial strength and highly efficient method for maintaining consensus on the ledger, capable of handling hundreds, or even thousands, of TpS, needs to be designed.
- And, most importantly, a satisfactory method for solving DvP has to be found.[5]

[3] Given that records of Bitcoin transactions are preserved in perpetuity, it might not be as good as believed for such activities.

[4] In real life, even movements from native Bitcoin addresses are performed with assistance of digital currency exchanges, such as Coinbase, which is orthogonal to the very idea of decentralization.

[5] Here, FBDC is coming into play. FBDCs, being fully fiat-backed tokens, reside on the ledger; since they are backed by the fiat currency at a one-to-one ratio, the corresponding DvP problem is solved naturally.

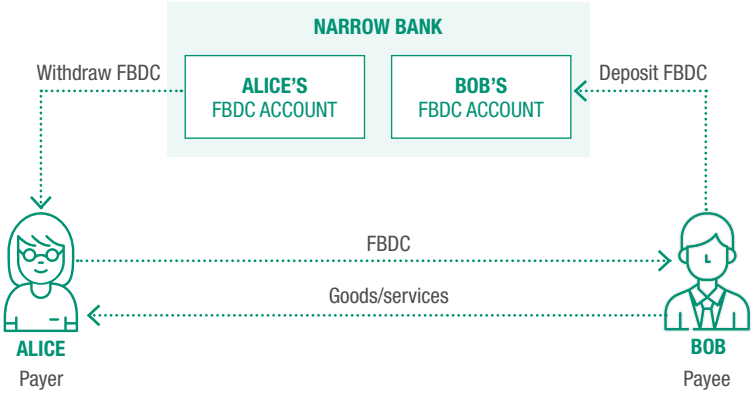
The validators (or notaries) responsible for the ledger integrity should be known in advance and licensed. They should be paid for their services a small fee, say a percentage of the transaction size they approve. This fee has to be denominated in FBDC, so that their interests and desires to maintain integrity of the ecosystem are properly aligned with their activities. In order to ensure Byzantine fault tolerance of the proposed setup, validators have to create their own versions of the ledger, and propose these to the rest of validators. Several rounds of voting take place until two-thirds majority is reached. In this regard, our approach is somewhat similar to the one used by Ripple [see Schwartz et al. (2014)], and can be viewed as a variation of the well-known Byzantine fault tolerant algorithms.

In order to provide an efficient and expedient transaction processing, individual notaries are assigned to particular subsets of all addresses. In this setup, a quorum verifies its portion of the ledger. And the full ledger is reconstructed out of these portions.

The only mechanism for injecting new coins in the distributed ledger is as follows. A participant has to have a conventional fiat account, either directly in the NB or with another commercial bank. They transfer the desired amount of fiat currency to the NB. The NB, in turn issues FBDC and transfers them from its public key address to the public key address provided by the participant. Thus, in effect, the participant becomes a shareholder in the NB rather than a depositor. Conversely, when a participant in the ledger wishes to receive fiat currency in exchange for their FBDCs, they transfer FBDCs from their public key to the public key of the bank, who, in turn, credit fiat currency either to the account on its own ledger or to a designated account in a different bank at a one-to-one ratio. Once a FBDC is born, it starts its journey from one address, represented by a public key, to the next. In this setup, the integrity of the distributed ledger is maintained by notaries.

In an alternative setup, coins are actually numbered and the list of numbers is maintained by the NB (although NB is unaware of which participant hold which number) in the blind-signature framework introduced by Chaum et al. (1990). Every time a coin changes hands, the new owner sends the number for checking by the NB, who compares it with the list of spent coins it maintains. If this particular coin has not been spent, it is retired, and a new coin with a new random number is issued to the designated owner. If the coin had been already spent, a transaction is rejected. The number is naturally blind-signed by the NB with its secret key, in order to prevent forgery and fraud.

Figure 4: A sketch of a transaction between Alice and Bob, in which Alice sends Bob 100 FBDCs



In Figure 4 we show a typical transaction between Alice and Bob, both having FBDC accounts.

We emphasize that the FBDC is a special case of the Digital Trade Coin (DTC), backed by a pool of real commodity assets, which is currently being developed at MIT, see Lipton and Pentland (2018) and Lipton et al. (2016).

3. ENTER A NARROW BANK

3.1 History

Modern banking originated in the High Middle Ages and blossomed during the Renaissance and the early modern period, mostly in the form of fractional reserve banking. From the beginning, fractional reserve banking firms were prone to collapse. For instance, in Florence the Bardi, Peruzzi, and Medici companies (to mention but a few) all failed.

Not surprisingly, the narrow banking idea was pursued by visionaries, financial reformers, and regulators for hundreds of years [see, e.g., Pennacchi (2012), Dittmer (2015), Roberds and Velde (2014), narrowbanking.org, and references therein]. From time to time, actual attempts to build a NB have been undertaken. For instance, in 1361 Venice's Senate prohibited lending out depositors' money, thus, in effect, making Venetian banks narrow. However, this prohibition was systematically circumvented, with associated bank failures to follow. In particular, the largest bank of

Pisano & Tiepolo failed in 1584, was converted into a state bank, and defaulted again in 1619. In 1609 the Bank of Amsterdam was chartered as an NB, but soon after started to lend its reserves in secret. As a result, in 1791 it failed and was taken over by the city.

Eventually banks, pursuing their own self-interests, became much more narrow than they were in the Renaissance, the early modern period, or are today. During the nineteenth century, British and American commercial banks followed the real bills doctrine and lent predominantly for short maturities. Bank loans mostly financed short-term working capital and provided trade credit, with maturities of two to three months, and were collateralized by borrower's personal wealth or the goods in transit [see Bodenhorn (2000) and Pennacchi (2012)].

In the twentieth century, however, encouraged by the creation of the Federal Reserve Bank in 1913, commercial banks drifted away from the real bills doctrine, started to lend for much longer maturities, established revolving lines of credit for some of their borrowers, and started to overemphasize their maturity transformation ability at the expense of prudence. The Great Depression of 1929 made banks' inability to meet their obligations successfully painfully obvious, which caused the idea of a NB to come to the fore.

In the U.K., NBs were advocated by Soddy (1926, 1933). In the U.S., a group of influential Chicago economists proposed a plan calling for the abolition of fractional reserve banks [see Knight et al. (1933), Hart (1935), Douglas et al. (1939), Fisher (1945)]. Their core proposals are summarized in Phillips (1996) as follows:

- Federal Reserve Banks should be owned by the government outright.
- Deposits of member banks should be completely guaranteed.
- Demands for payments by depositors should be satisfied by issuing Federal Reserve Notes as legal tender.
- The gold standard should be suspended.
- The assets of all member banks should be liquidated and all existing banks dissolved.
- New NBs accepting only demand deposits subject to 100% reserve requirement in cash and deposits with the Fed should be created.
- Investment trusts handling saving deposits should be created.
- Existing banking institutions should operate under Federal Reserve supervision until they are dissolved and new banks are created.

Although a practical conversion of fractional reserve banks into NBs was rejected in the forties under enormous political pressure from fractional reserve banks, the idea has always stayed close to the surface, and gained considerable momentum during and after the S&L crisis in the 1980s and 1990s [see, e.g., Friedman (1959), Tobin (1986), Litan (1987), Bryan (1991), Burnham (1991), Gorton and Pennacchi (1993), Huber and Robertson (2000), Kobayakawa and Nakamura (2000), Al-Jarhi (2004), Garcia et al. (2004)]. Not surprisingly, it became extremely popular again during and after the GFC [see, e.g., Kay (2010), Kotlikoff (2010), Phillips and Roselli (2011), Benes and Kumhof (2012), Chamley et al. (2012), Pennacchi (2012), van Dixhoorn (2013), Admati and Hellwig (2014), Cochrane (2014), Dittmer (2015), Garratt et al. (2015), McMillan (2015)].

3.2 A bank that cannot default

The main characteristic of a NB is its assets mix, which includes solely marketable low-risk securities and central bank cash in the amount exceeding its deposit base. As a result, such a bank can only be affected by operational failures, which can be minimized, but not eliminated, by using state-of-the-art technology, thus providing a maximally safe payment system. Accordingly, NB deposits would be equivalent to currency, thus abolishing the need for deposit insurance with all its perverse effects on the system as a whole, not to mention the associated moral hazards.

It is clear that the only way to keep a one-to-one parity between the fiat currency and digital tokens is to keep the exact amount of the fiat in escrow. However, you cannot put the requisite amount in a bank and expect it to be safe at all times, unless this bank is specially designed, or else you can open an account directly at the central bank. Indeed, bank depositors are junior unsecured creditors of a bank, so if the bank were to default, they cannot expect their deposits to stay intact. Even if a significant portion of these deposits can be recovered, the money will not be available until the bankruptcy issues are resolved, which can take a very long time. At the same time, a central bank, while happy to accommodate licensed banking institutions and a small selected group of trusted non-banking financial firms, such as central clearing counterparties,

cannot, and will not, allow a wider range of corporate or individual participants (particularly, if they wish to be anonymous) to have account with them. This is for a variety of reasons, including, but not limited to, being unable to solve the KYC/AML problem, not to mention potential political complications.

Thus, we need to build a bank, which cannot default, at least due to market and liquidity risks. One needs to be cognizant of the fact that, regardless of the amount of effort, it is not possible to build a bank impervious to operational risks, although proper design can minimize them to an acceptable degree.

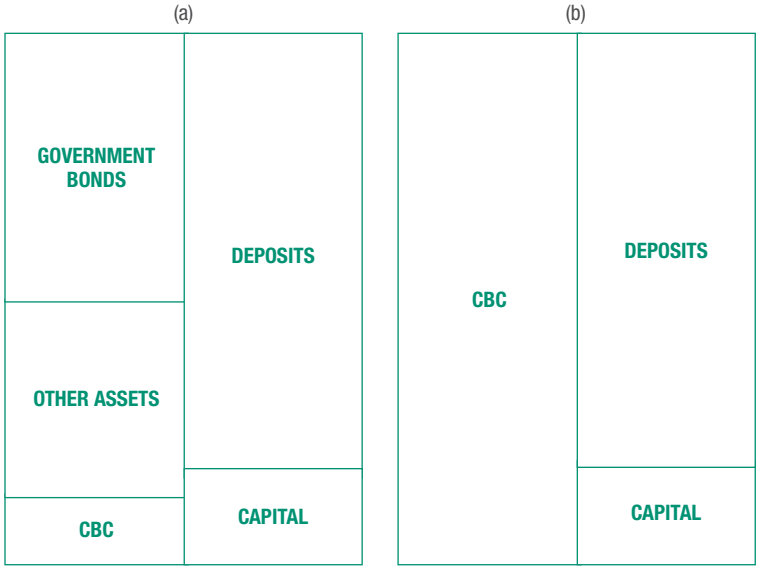
3.3 Types of NBs

Several approaches for designing a NB have been summarized in Pennacchi (2012):

- 100 percent reserve bank (C-PeRB): Assets – central bank reserves and currency; liabilities – demandable deposits and shareholder equity. Depending on the circumstances, these deposits can be either noninterest-bearing, or interest paying, or interest charging. The latter set-up might be necessary if interest rate paid by central bank is negative. C-PeRB is financed by a combination of deposits (debt) and shareholders' equity.
- Treasury money market mutual fund (TMMMF): Assets – Treasury bills or repurchase agreements collateralized by Treasury bills; liabilities – demandable equity shares having a proportional claim on the assets. TMMMF is financed solely by equity.
- Prime money market mutual fund (PMMMF): Assets – short term Federal agency securities, short-term bank certificates of deposits, bankers' acceptances, highly rated commercial paper, and repurchase agreements backed by low-risk collateral; liabilities – demandable equity shares having a proportional claim on the assets. As before, PMMMF is financed solely by equity.
- Collateralized demand deposit bank (CDDB): Assets – low-credit- and interest-rate-risk money market instruments, which are fully (over)-collateralized; liabilities – demandable deposits that have a secured claim on the collateral.
- Utility bank (UB): UB is similar to a CDDB, except for the fact that collateral can include retail loans in addition to money market instruments;

Putting aside operational risks inherent in the banking business, the reliability of a NB varies from completely stable (C-PeRB), to stable under most plausible circumstances (UB).

Figure 5: Balance sheets of a fractional reserve bank and an NB



The difference between balance sheets of a fractional reserve bank and a NB is shown in Figure 5.

3.4 The time for a NB is now

Whilst running a NB is relatively easy from a market perspective, and the required capital for doing so is comparatively small (under current Basel regulations its size is determined by leverage alone), it naturally has to possess bullet-proof security and reliability. These requirements can be met by judiciously building the corresponding ledger software and hardware. Of course, in addition to pure operational aspects, the NB has to satisfy the KYC/AML requirements. It is clear that a liberal usage of "artificial intelligence," "machine learning," and "big data analytics" is necessary to accomplish this task efficiently. In this regard, TRUST::DATA, a new framework for identity and data sharing currently being developed at MIT, is particularly promising [see Hardjono et al. (2016)].

There is a perennial question of profitability of a NB. Whilst a fractional reserve bank earns its living first and foremost via the "net interest margin" (NIM), i.e., the difference between the interest it charges its borrowers and interest it pays its depositors, a NB seemingly is deprived of this all important source income. However, this is only partially true, since at present some central banks, including the Federal Reserve, do pay substantial interest on excess deposits. Besides, NBs can earn interest on securities, charge reasonable fees

for transaction services, etc. While their operational margins are certainly low (by yesteryear standards), so are their capital requirements, operating costs (due to an efficient infrastructure), and regulatory burdens. Thus, NBs could generate competitive returns on equity, which are very favorably compared to the ones generated by their fractional reserve cousins. The quote from Friedman (1959) captures the essence of the problem: "I shall depart from the original 'Chicago Plan of Banking Reform' in only one respect, though one that I think is of great importance. I shall urge that interest be paid on the 100% reserves. This step will both improve the economic results yielded by the 100% reserve system, and, also, as a necessary consequence, render the system less subject to the difficulties of avoidance that were the bug-a-boo of the earlier proposals. ... This problem of how to set the rate of interest is another issue that I feel most uncertain about and that requires more attention than I have given to it."

If NBs in different jurisdictions organize themselves as a network of sister banks, they can earn substantial (but fair) transactional fees on foreign exchange transactions.

In principle, NBs can be affiliated with lending organizations with uninsured funding, the so-called lending affiliates. In view of this fact, lending facilities can be left to their own devices and be regulated by market forces.

It is clear that the adoption of narrow banking in its entirety would require a massive transformation of the financial ecosystem and should not be undertaken until numerous and nuanced questions dealing with the pros and cons of such a transformation are answered in sufficient detail. While we list some of the pros and cons below, we are interested in a less ambitious project – that is an introduction of an NB, which would coexist with fractional reserve banks, rather than supplant them completely. An interesting analogy jumps to mind – currently electric cars (NBs), coexist with conventional gasoline cars (fractional reserve banks). While in the long run electric cars are likely to prevail over gasoline cars, in the short run they can peacefully cohabit. In order to avoid academic discussions related to the transformation of the banking system from the fractional reserve to the narrow setup, we advocate creation of a few NBs as needed for achieving our specific goals. We anticipate coexistence of fractional reserve and NBs for a long time to come.

4. PROS AND CONS OF A NB

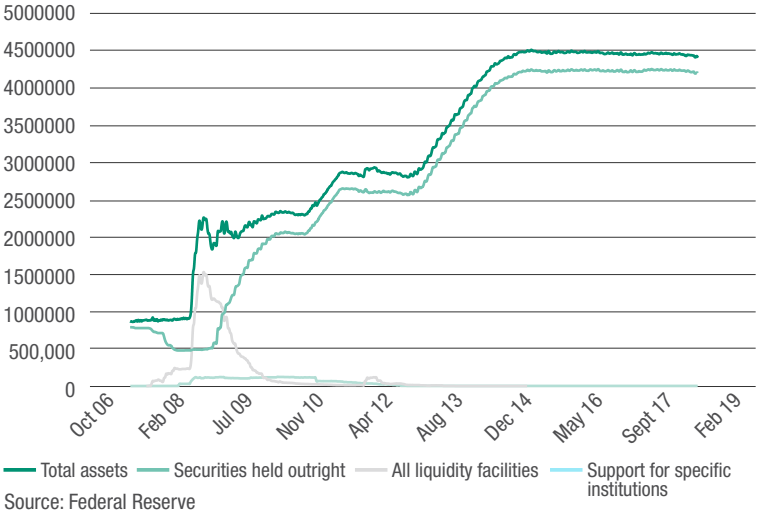
4.1 Pros

There are many leading economists who advocate narrow banking because some of its benefits are self-evident. First, by construction, and in contrast to fractional reserve banks, assets and liabilities of NBs are perfectly aligned, so that conventional stabilization mechanisms such as deposit insurance, discount window lending, rigorous regulation and control of balance sheet, without which fractional reserve banks cannot exist, are simply not necessary. We emphasize, however, that other types of regulation are certainly needed, not least because NBs, like any other organizations, are subject to operational risks, particularly from electronic attacks.

"Fortunately, remarkable technological breakthroughs – mostly related to cryptocurrencies, distributed ledgers, and related concepts – simultaneously focused attention of key decision-makers and technical experts on the glaring need for transforming the financial infrastructure, and, at the same time, indicated how such a transformation can be accomplished."

Second, since lending is performed by non-banking institutions on an uninsured basis, governmental interference in bank lending and other activities can

Figure 6: Assets of the Federal Reserve Bank



Source: Federal Reserve

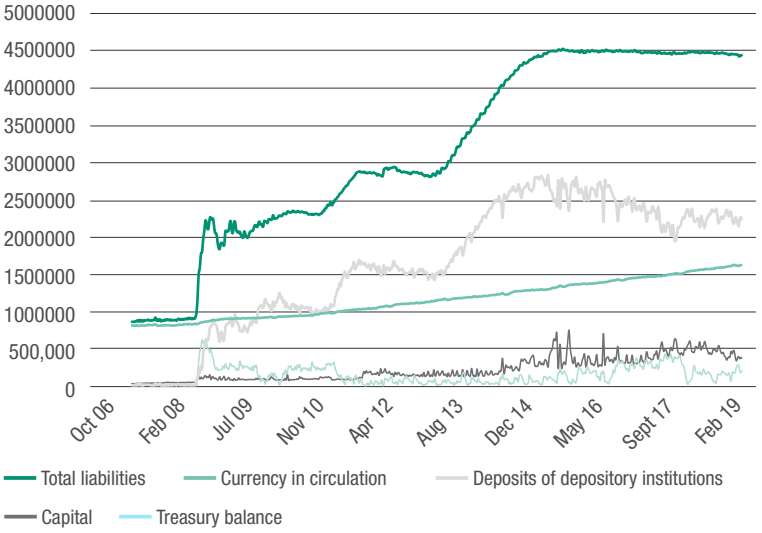
be dramatically reduced, if not completely eliminated. Third, deposit insurance can be reduced in size and eventually phased out.

4.2 Cons

Needless to say, narrow banking is not without its detractors. Some economists argue that NB will not be a silver bullet needed to kill financial instability, particularly because lending affiliates would suffer from the same issues as fractional reserve banks. Although it is true to some extent, it is clear that narrow banking can serve as a cornerstone of a stable and reliable payment system, capable of operating on its own even under the most extreme conditions, so that the pressure on the financial ecosystem as a whole would be significantly less compared to fractional reserve banking. To attract investors, lending affiliates would have to maintain their own strong capital cushion and look for long-term financing opportunities. Still, these measures in and by themselves might not be sufficient to ensure the financial stability under all circumstances, so that the "lender of last resort" in the form of a central bank would still have to be present in the system. Such a bank will provide required liquidity to uninsured lenders including affiliates of NBs against illiquid, but sound, collateral, thus avoiding a systemic credit collapse. This is to be compared with the current setup, where financial authorities support private banks through deposit insurance, access to the discount window, and implicit government guarantees.

Specifically, Miles (2001) argues that separation of deposit taking and lending would result in elevated agency costs and reduce stability of the supply of lending. In all likelihood, this is not going to happen since lenders would become much more efficient to survive without a cushion provided by depositors. Bossone (2002) emphasizes that benefits of NB in terms of financial stability are much smaller than its drawbacks associated with cutting the link between bank money and economic activity and creating "market incompleteness." He thinks that this void will be filled by financial firms, whose operations will be as risky as the ones conducted by fractional reserve banks, so that overall stability of the financial ecosystem will not improve. Most interestingly from our standpoint, Bossone (2002) is not opposed to voluntary creation of NBs, or segregated NB subsidiaries within existing bank holding companies.

Figure 7: Liabilities of the Federal Reserve Bank



Source: Federal Reserve
Notes: Excess reserves kept by commercial banks increased enormously since 2008

The other danger is the risk of flight to quality from fractional reserve banks to NBs during the times of financial instability, i.e., precisely when the former can least afford to lose their liquidity. This danger is not as acute as it might sound, because the actual amount of liquidity NBs can absorb is limited by their capital size.

5. NBS AS PART OF THE FINANCIAL ECOSYSTEM

5.1 Current trends in banks' behavior

In the build-up to the GFC, banks tried to stay as leveraged as possible, by simultaneously reducing their capital ratio and choosing progressively riskier asset mix. However, after 2008, their group behavior changed dramatically. The balance sheet of the Federal Reserve is shown in Figures 6 and 7. Comparison of these figures shows that the asset and liability mix of the banking sector underwent a dramatic transformation after the GFC. One of the most striking aspects of this change is the precipitous increase in excess reserves depositary institutions keep with the Federal Reserve. We are observing interesting and somewhat perplexing developments: until the onset of the GFC, central banks were run as NBs, and commercial banks were run as fractional reserve banks, while after the crisis the situation flipped, although not completely. This fact shows that banks prefer to keep a considerable cash cushion, partly because they put an extra premium

on maintaining high liquidity, and partly due to lack of demand for loans. Besides, attractive interest rate paid by the Federal Reserve on the excess deposits is clearly an additional motivation.

In view of the above, it is clear that building a NB cannot and should not upend the overall balance of the banking ecosystem, since it is pretty much aligned with prevailing trends anyway.

5.2 What can an NB do for you?

A properly designed NB is a natural repository of funds for those who highly value their funds' stability (either by inclination, such as wealthy individuals and organizations, or by necessity, such as central clearing counterparties). It is also a natural emitter of FBDC. In addition, such a bank can do many other things. For instance, it can be used to hold non-operational deposits, which conventional commercial banks do not want and cannot hold at a profit. Besides, it is a custodian for initial margins (IM) supplied by investment banks as part of their regular over-the-counter derivatives business. These funds are naturally paid via FBDC and are kept safe by construction. Moreover, if so desired, the NB, being a neutral custodian, can provide value-added services, such as calculating the size of the required collateral and administering its allocation. Besides, a NB can be a very useful source of digital identity.

5.3 Lending affiliates – credit money creators of the future

If banking institutions all become narrow, then credit creation will be performed by lending affiliates and other lenders, for instance, mutual funds or hedge funds. In fact, after the GFC, a considerable portion of credit is issued by non-banks, while many banks keep massive excess reserves with central banks, thus becoming de-facto more narrow. By reorganizing themselves into transaction-oriented NBs and lending affiliates, fractional reserve banks can become much more cost-efficient, nimble, and stable.

By construction, NBs offer their depositors a high level of safety, handle regulatory burden with relative ease, require a low capital cushion, derive a stable and considerable flow of income from their transactional activities, and benefit from the interest paid on bank reserves by central banks. Transactional cash flow can be increased manifold if foreign exchange and, especially, cryptocurrency issuance are included into the mix. At the same time, since NBs require very limited capital cushion, which is needed to satisfy leverage ratio constraints and cover operational risks, they can offer very attractive return on equity (ROE) to their investors. Recall that a non-risk-based leverage ratio is calculated by dividing Tier-1 capital by the bank's average total consolidated assets, which, for NBs, boil down to central banks reserves and short-term government paper. Under Basel III rules, banks have to maintain a leverage ratio in excess of 3%.

Given the simplicity of their balance sheet and efficiency of the state-of-the-art IT systems, NBs can use technological advancements, such as distributed ledgers and blockchain, to provide excellent transactional banking services and successfully complete with transactionally-oriented fintech startups [see Lipton (2016a), Lipton et al. (2016), He et al. (2017), Powell (2017), among others].

At the same time, uninsured lending affiliates of NBs, unencumbered by the requirement to provide utility-like transactional services, can better serve the needs of the real economy, by offering traditional as well as innovative credit financial products. Given that lending affiliates would not have cheap sources of funding in the form of deposits, they would have to maintain healthy capital cushions and choose the quality of assets aligned with their risk appetite, in order to attract savings and other forms of funding from investors. Lending affiliates would be stratified depending on the level of their speculative activities. Denuded of all amenities related to deposit insurance, lending affiliates will have their own skin in the game, and be open to scrutiny by their investors.

Thus, splitting fractional reserve banks into NBs and lending affiliates would increase investment value of both, much like nuclear fission releases enormous energy in nature.

5.4 Limited impact of narrow banks on the ecosystem at large

Even though an NB by construction is impervious to market and liquidity shocks, it can suffer from operational risks. Hence, it requires capital cushion. The size of this cushion is determined by the leverage ratio, and is of the order 3%-4% of its assets.

Thus, the size of the available capital effectively limits the amount of central bank money a NB can attract from fractional reserve banks. As a result, potential systemic impact of such an institution on the financial system as a whole is limited. Besides, since a NB does not lend its funds, it is unable to create money “out of thin air,” so from this angle, its impact is limited too.

Yet, such a bank would have a great impact in other ways. First and foremost, it will create an honest competition in the banking ecosystem and will force conventional banks to pay a fair interest to their depositors. Second, it would make FBDC expansion above and beyond its original narrow base a reality. Finally, for the first time in recent history, such a bank would provide a venue for both retail and institutional depositors who are particularly concerned about availability and stability of their deposits even under the most extreme conditions. Among the institutional depositors, central clearing counterparties are the primary candidates, given that they have all kind of negative externalities including the fact that some of their largest clearing members are, at the same time, their bankers. Thus, a potential default of a clearing member can cause a double loss for such a CCP.

NB, being a radical departure from the familiar financial setup, naturally raises numerous questions of monetary policy, particularly regarding the manner of money creation and who should be responsible for it. The main issue is that to a large extend money will be created or destroyed by central banks, which would have to exercise preternatural abilities to do so properly. Money creation along these lines would be a de facto tool of central planning. Given that central planning is next to impossible to execute efficiently, the dangers can outweigh the benefits. The behavior of credit markets would be affected in a very profound way, since banks will no longer be natural sources of credit. All these effects have to be analyzed in detail before narrow banking is implemented in its entirety.

6. CBDC VERSUS FBDC

In principle, distributed ledgers can potentially become a truly transformative force by making central bank digital currency (CBDC) a reality, in a dramatic departure from the past. A variety of viewpoints on this subject, some of which are mutually exclusive and contradictory, can be found in Ali et al. (2014), Andalfatto (2015), Barrdear and Kumhof (2016), Broadbent (2016), Danezis and Meiklejohn (2015), Fung and Halaburda (2016), Koning (2016), Lipton (2016b), Bordo and Levin (2017), Dyson and Hodgson (2016), Mersch (2017), Scorer (2017), among many others.

If central banks start to issue CBDC, they can not only abandon physical cash in favor of its electronic equivalent, as is advocated in Rogoff (2016), but, eventually, retire a substantial portion of the government debt in its favor. This would be a very impactful development for society at large. Taken to its logical limit, CBDC can eliminate fractional banking raison d’être and dramatically improve financial ecosystem resilience, by allowing economic agents to have accounts at the central bank directly. As a result, these

will dramatically reduce the ability of the banking sector to create money "out of thin air" and transfer this all-important function to central banks. However, central banks are not equipped to address the large-scale KYC/AML problem, which they would have to solve if they open their balance sheets to a large portion of economic agents, rather than licensed banks and selected financial institutions alone. While developments in this direction are inevitable, their timing and magnitude cannot be ascertained at present.

Realistically, we do not expect central banks' balance sheet to be open to all economic agents. Accordingly, we think that FBDC, being a private coin, is a much more convenient solution to digitization of the fiat currency than CBDC. Issued by a purpose-built NB, FBDC will be as reliable as fiat. At the same time, the corresponding bank can satisfactorily solve the KYC/AML problem and navigate the complicated political landscape. Moreover, NBs, organized as a network of sister banks incorporated in different jurisdictions, can simplify and cheapen foreign exchange transactions.

7. DIGITAL IDENTITY AND KYC/AML

With the emergence of blockchain and DLT, and their usage for cryptocurrencies, the question of digital identity in the context of KYC/AML has come to the foreground. A major shortcoming in current identity systems on the internet is the lack of privacy with respect to transactions performed using these identities. This deficiency is also true in the context of blockchain-based currencies, such as Bitcoin, namely the disclosure of identities through the reverse engineering and analytics of the public-keys used in transactions recorded on the blockchain.

We believe a new breed to "crypto-identities" may need to be devised that not only provides transaction confidentiality, but more importantly exhibit the features necessary to make it compliant to KYC/AML regulations. These crypto-identities must be based on and derive from the appropriate combination of highly private and accurate personal data, and must yield truthful assertions or claims regarding the owner relevant to the KYC/AML requirements. Additionally, for transaction confidentiality, these identities must be conditionally anonymous-verifiable, meaning that the identities must seemingly be anonymous to non-participants and be reversible by KYC/AML processes. In this way, a chain of provenance (or chain of verifiability) can be established for a given digital identity from the transaction on the blockchain to the legal owners of the digital identity.

The area blockchain and DLT is currently still nascent, and additional infrastructure technologies will be needed in order for the full benefits of blockchains to be realized in a transformative manner in connection to digital identities. The report by Hardjono and Maler (2018) provides a broad industry review of identity technology and the relevance of blockchain to identity management.

8. MORAL HAZARD

One of the greatest hazards of a widely used digital currency is enabling a repressive surveillance state. If the government can track all of its citizens' payments, then they can exert unprecedented control over their lives. Nor is this situation just some science fiction fantasy; in parts of Northern China virtually all payments – for transportation, food, entertainment, communication, everything – are logged by just two companies, both of whom collaborate closely and share data with the government.

To avoid this situation, small financial transactions, such as currently performed with cash, must be anonymous. Exceptions to this anonymity should be few and far between. For instance, in serious criminal investigations or similar situations, where there is an overriding social imperative, society may decide that it should be possible to override this anonymity using carefully vetted and expensive methods such as legal court orders.

Fortunately, there are a range of cryptographic methods to enforce levels of anonymity ranging from technologies that allow complete unbreakable anonymity, to methods that provide anonymity for payers but not for sellers, to frameworks that provide anonymity except for court orders. For instance, a narrow bank can follow the Chaumian scheme and issue numbered and blind signed currency units onto a distributed ledger, whose trust is maintained either by designated notaries or by the bank itself. KYC/AML requirements could be limited to large deposits or withdrawals, much as cash transactions are today.

9. CONCLUSION

In this document, we have outlined an efficient framework, which can be used in order to extend the domain of applicability of the FBDC from an initial group of bank sponsors to a much wider group of potential users including SMEs. We have argued that a purpose-built NB is necessary (and, hopefully, sufficient) to achieve this goal. Not only can it be used to securely hold collateral, but also to solve the all-important KYC/AML problem. The FBDC, being a stable cryptocurrency, can facilitate both domestic and foreign trade and offer numerous possibilities for streamlining and facilitating commercial and retail transactions.

References

Al-Jarhi, M., 2004, "Remedy for banking crises: What Chicago and Islam have in common: A comment," Islamic Economic Studies 11:2, 23-42

Admati, A. and M. Hellwig, 2014, The bankers' new clothes: what's wrong with banking and what to do about it, Princeton University Press

Ali, R., J. Barrdear, R. Clews, and J. Southgate, 2014, "The economics of digital currencies," Bank of England Quarterly Bulletin 54:3, 276-286

Andolfatto, D., 2015, "Fedcoin: on the desirability of a government cryptocurrency," Web Log post

Barrdear, J. and M. Kumhof, 2016, "The macroeconomics of central bank issued digital currencies," Bank of England Working Paper No. 605

Benes, J. and M. Kumhof, 2012, "The Chicago plan revisited," International Monetary Fund, WP/12/202

Bodenhorn, H., 2000, A history of banking in Antebellum America: financial markets and economic development in an Era of Nation-Building, Cambridge University Press

Bordo, M. D. and A. T. Levin, 2017, "Central bank digital currency and the future of monetary policy," National Bureau of Economic Research No. w23711

Bossone, B., 2002, "Should banks be narrowed? An evaluation of a plan to reduce financial instability," Public policy brief No. 69, Jerome Levy Economics Institute of Bard College

Broadbent, B., 2016, "Central banks and digital currencies," Speech at London School of Economics

Bryan, L., 1991, "Core banking," The McKinsey Quarterly 1, 61-74

Burnham, J. B., 1991, "Deposit insurance: the case for the narrow bank," Regulation 14, 35-43

Buterin, V., 2013, "What proof of stake is and why it matters," Bitcoin Magazine, August 26

Castro, M. and B. Liskov, 1999, "Practical Byzantine fault tolerance," OSDI 99, 173-186

Chamley, C., L. J. Kotlikoff, and H. Polemarchakis, 2012. "Limited-purpose banking-moving from "trust me" to "show me" banking," American Economic Review 102:3, 113-119

Chaum, D., A. Fiat, and M. Naor, 1990, "Untraceable electronic cash," in Proceedings on advances in cryptology, Springer-Verlag, 319-327

Cochrane, J. H., 2014, "Toward a run-free financial system," SSRN 2425883

Danezis, G. and S. Meiklejohn, 2015, "Centrally banked cryptocurrencies," arXiv preprint arXiv:1505.06895

Dittmer, K., 2015, "100 percent reserve banking: a critical review of green perspectives," Ecological Economics 109, 9-16

van Dixhoorn, C., 2013, "Full reserve banking: an analysis of four monetary reform plans," Sustainable Finance Lab

Douglas, P. H., I. Fisher, F. D. Graham, E. J. Hamilton, W. I. King, and C. R. Whittlesey, 1939, A program for monetary reform

Dyson, B., and G. Hodgson, 2016, "Digital cash. Why central banks should start issuing electronic money," Positive Money, http://bit.ly/2Fn26Uf

Fisher, I., 1945, 100% money; designed to keep checking banks 100% liquid; to prevent inflation and deflation; largely to cure or prevent depressions: and to wipe out much of the national debt, City Printing Company

Friedman, M., 1959, A program for monetary stability, Fordham University Press

Fung, B. S., and H. Halaburda, 2016, "Central bank digital currencies: a framework for assessing why and how," Bank of Canada Staff Discussion Paper no. 22

Garcia, V. F., V. F. Cibils, and R. Maino, 2004, "Remedy for banking crises: what Chicago and Islam have in common," Islamic Economic Studies 11:2, 1-22

Garratt, R., A. Martin, J. McAndrews, and E. Nosal, 2015, "Segregated balance accounts," Federal Reserve Bank of New York, Staff Report no. 730

Gorton, G., and G. Pennacchi, 1993, "Money market funds and finance companies: are they the banks of the future?" in Structural change in banking, Irwin Publishing

Hardjono, T., and E. Maler (eds.), 2018, Report from the Blockchain and Smart Contracts Discussion Group to the Kantara Initiative, Kantara Initiative

Hardjono, T., D. Shrier, and A. Pentland, 2016, "TRUST:: DATA: a new framework for identity and data sharing," Visionary Future LLC.

Hart, A. G., 1935, "The "Chicago Plan" of banking reform: a proposal for making monetary management effective in the United States," Review of Economic Studies 2:2, 104-116

He, M. D., M. R. B. Leckow, M. V. Haksar, M. T. M. Griffoli, N. Jenkinson, M. M. Kashima, T. Khiaonarong, M. C. Rochon, and H. Tourpe, 2017, "Fintech and financial services: initial considerations," International Monetary Fund

Huber, J., and J. Robertson, 2000, Creating new money: a monetary reform for the information age, New Economics Foundation

Kay, J., 2010, "Should we have narrow banking," London School of Economics

Knight, F., G. Cox, A. Director, P. Douglas, A. Hart, L. Mints, H. Schultz, and H. Simons, 1933, Memorandum on banking reform, Franklin D. Roosevelt Presidential Library, President's Personal File, 431

Kobayakawa, S., and H. Nakamura, 2000, "A theoretical analysis of narrow banking proposals," Monetary and Economic Studies 18:1, 105-118

Koning, J., 2016, "Fedcoin: a central bank-issued cryptocurrency," R3 Report 15

Kotlikoff, L. J., 2010, Jimmy Stewart is dead: ending the world's ongoing financial plague with limited purpose banking, John Wiley & Sons

Lamport, L., R. Shostak, and M. Pease, 1982, "The Byzantine generals problem," ACM Transactions on Programming Languages and Systems (TOPLAS) 4:3, 382-401

Lipton, A., 2016a, "Banks must embrace their digital destiny," Risk Magazine 29, 8

Lipton, A., 2016b, "The decline of the cash empire," Risk Magazine 29, 53-53

Lipton, A., T. Hardjono, and A. Pentland, 2018, "Digital trade coin (DTC): towards a more stable digital currency," Submitted for publication

Lipton, A., and A. Pentland, 2018, "Breaking the bank," Scientific American 318:1, 26-31

Lipton, A., D. Shrier, and A. Pentland, 2016, "Digital banking manifesto: the end of banks?" Massachusetts Institute of Technology

Litan, R. E., 1987, What should banks do? Brookings Inst Press

Mersch, Y., 2017, "Digital base money: an assessment from the ECB's perspective," Speech at the Farewell ceremony for Pentti Hakkarainen, Deputy Governor of Suomen Pankki--Finlands Bank, Helsinki

Micali, S., 2016, "ALGORAND: the efficient and democratic ledger," arXiv preprint arXiv:1607.01341

Miles, W., 2001, "Can narrow banking provide a substitute for depository intermediaries?" Department of Economics, Wichita State University, 30

McMillan, J., 2015, The end of banking: money, credit, and the digital revolution, BookBaby

Nakamoto, S., 2008, Bitcoin: a peer-to-peer electronic cash system

NarrowBanking, www.narrowbanking.org

Norman, B., R. Shaw, and G. Speight, 2011, "The history of interbank settlement arrangements: exploring central banks' role in the payment system," Bank of England Working Paper No. 412

Pennacchi, G., 2012, "Narrow banking," Annual Review of Financial Economics 4:1, 141-159

Phillips, R. J., 1996, "The `Chicago Plan' and new deal banking reform," in Stability in the financial system, Palgrave Macmillan

Phillips, R. J., and A. Roselli, 2011, "How to avoid the next taxpayer bailout of the financial system: the narrow banking proposal," in Financial market regulation, Springer

Powell, J., 2017, "Innovation, technology, and the payments system," Speech: blockchain: the future of finance and capital markets

Roberds, W., and F. R. Velde, 2014, Early public banks

Rogoff, K., 2016, The curse of cash, Princeton

Schwartz, D., N. Youngs, and A. Britto, 2014, "The Ripple protocol consensus algorithm," Ripple Labs Inc white paper, 5

Scorer, S., 2017, "Central bank digital currency: DLT or not DLT? That is the question." BankUnderground http://bit.ly/2FiUV3n

Soddy, F., 1926, Wealth, virtual wealth, and debt, 2nd American Edition, 1933, E.P. Dutton

Soddy, F., 1933), Money versus man, Dutton

Tobin, J., 1986, "Financial innovation and deregulation in perspective," Cowles Foundation for Research in Economics, Yale University

Quantitative investing and the limits of (deep) learning from financial data

J. B. HEATON | Managing Member, Conjecture LLC

ABSTRACT

The idea of quantitative investing – using robust computing power and algorithms to trade securities – inspires both awe and fear. Reality is less exciting. With a tiny handful of exceptions, most quant funds have been unimpressive. I explore some limits of quantitative investment, with a focus on the promise – or lack thereof – of techniques from deep learning and artificial intelligence. These limitations help explain the disappointing performance of many quant strategies and cast doubt on the promise of artificial intelligence techniques for improving returns. The main problem is that financial market data is unlike the data that machine learning works well on computer vision, speech recognition, and natural language processing. While deep learning and artificial intelligence are changing the world in many ways, they are unlikely to generate fortunes for investors, who will continue to remain best-served by inexpensive and passive index products that themselves will be augmented by machine learning techniques to drive costs even lower.

1. INTRODUCTION

Quantitative investing – using robust computing power and algorithms to trade securities – inspires awe and fear. Awe arises from the idea that the use of mathematics, statistics, and computer learning might be the closest thing possible to the real-world Philosopher's Stone. Fear comes from the worry that traders using computer algorithms to trade securities – often at the "high frequency" of microseconds and faster – can disrupt capital markets and delink prices from fundamentals. Investors have responded more to the awe than to the fear. Investment industry sources estimate that quant funds managed more than U.S.$1 trillion by the beginning of 2018. The strategies include everything from relatively simple regression-based "factor models" from firms like AQR (founded by fellow finance PhDs from the University of Chicago) to highly-complex computer learning models employing the latest ideas from artificial intelligence, including the deep learning models I refer to in my title.

The reality is less impressive than awe or fear suggest. One of the biggest fund implosions of all time was Long-Term Capital, a quant fund run by former Salomon Brothers proprietary traders and Nobel Prize winners Robert Merton and Myron Scholes. More recently, BlueTrend, a Geneva-based fund and Winton Capital, a large London-based fund, have had unimpressive years, as has AHL, one of the main funds of the Man Group, and Aspect Capital, another large quant fund. With only a handful of exceptions – most notably Renaissance Technologies, the hedge fund that mathematician James Simons founded in 1982 – most quant funds do not have consistently impressive performance. Consider Citadel LLC, a huge hedge fund headquartered in Chicago. The hedge fund returned only about 13% in 2017, short of the S&P 500 Index's 20% gain, a return available quite inexpensively to anyone with the money to open an account at Vanguard Group. The explanation for the difference in returns is not lower risk. Citadel fell nearly 60% in 2008, far more than the S&P 500 index. More recently, when market turmoil hit in early February 2018, some of the biggest names in quant fell hard again, including Winton and AHL, and a large quant fund managed by Lynx Asset Management.

In this article, I explore some limits of quantitative investing, with a focus on the promise – or lack thereof – of techniques from deep learning and artificial intelligence more generally. As deep learning – a subset of machine/computer learning – has achieved more and more success in image and speech recognition, product recommendations, and self-driving vehicles, hope has escalated that the techniques allowing advances in other domains will pay off for quant investors as well. Deep learning is a form of machine learning, the use of data to train a model to make predictions from new data. Recent advances in deep learning have dramatically improved the ability of computers to recognize and label images, recognize and translate speech, and play games of skill, in each case sometimes at better than human-level performance. In these applications of deep learning, the goal is to train a computer to perform, even better, certain tasks – such as recognizing the content of an image – that a human usually is able to do quite well.

Financial markets present far different problems than those presented in computer vision, speech recognition, and natural language processing. Unlike recognizing an image or responding appropriately to verbal requests, humans have no innate ability to, for example, select a stock that is likely to perform well in some future period. These limitations are related to the disappointing performance of many quant strategies and cast doubt on the promise of artificial intelligence techniques for improving matters.

2. STATISTICAL MODELING, MACHINE LEARNING, AND DEEP LEARNING

The idea behind many quant strategies is that some variable of interest – say, a move in a security price – can be modeled as some function of available data. The function of the data could be something as simple as a factor regression model or something as complex as a many-layered deep learning computation. We can conceive of the problem as approximating for the variable of interest some existent, but unknown, true function of the data. The problem is to use available data to estimate the relationship or "train the model" and then test the model on new data or data that we set aside from the initial estimation. The basic tools of quant strategies are mathematics (including optimization, probability, information theory, and statistics) and numerical methods using computers.

The problem is that financial markets generate data that is not like the data on which machine learning works well. Machine learning is incredibly powerful with data patterns that are stable. A vision system may take a while to learn all the ways to recognize a dog in an image, since there are so many different angles from which to capture the dog and so many objects that might interfere with the dog's image (e.g., the

dog might be partially hidden by a fire hydrant!). But a dog is a dog and remains so, and the algorithm can learn to distinguish dogs of different types from cats and cacti. The same is true of chess patterns, the game of Go, speech recognition, you name it. The computer is allowed to train on data and learn relationships that remain relatively stable and eventually gets it right.

Financial market data is different. First, the algorithm may identify a relationship that does not actually exist. "Signals" in financial markets come with enormous amounts of "noise." A quant system may falsely identify a signal that does not actually exist or may overestimate the effect of an actual signal. Second, even if the relationship did exist at one time in the data, it may disappear quickly. The behavior causing the relationship might change as investors learn something that alters their expectations. Alternatively, arbitrage – the actions of the quants themselves – might cause the signals to disappear as investors who see the relationship compete them away. Anyone who has taken an introductory economics course has encountered the model of perfect competition where buyers and sellers are equally well informed and there is no market power. Quants are all trying to hire the same kinds of people, educated at the same institutions, and trained in the same methodologies. As a result, quant trading is largely (Renaissance Technologies excepted?) a commodity business and competition to exploit signals looks a lot like competition to buy and sell identically-graded wheat. Third, if a relationship does not disappear – that is, a signal continues to "work" – the relationship may indicate the presence of a risk-return tradeoff, not an arbitrage opportunity. If a relationship that everyone can see continues to exist, then – like the fact that stocks on average return more than U.S. treasuries – it is more likely that the return is compensation for real risk. Indeed, the marketing genius of firms like Dimensional Fund Advisors and AQR is that they implemented simple academic factor model-based investing as if it offered superior risk-adjusted performance when that superior performance is probably just compensation for economic risk that occasionally bites the investors who bear it.

3. FOOLING THE MACHINE

A human – though perhaps quite fallible – is often able to notice or "get the feeling," through some mechanism of intuition we do not yet understand, that something is not quite right. This can allow a human trader to put the brakes on a strategy. Computers are not always good at this. Sometimes this is what allows algorithms to make better decisions, as they are not prone to wrong intuitions. But sometimes intuitions are right. Livermore (1940), the famed speculator of the early 20th century, writes: "A speculator of great genius once told me: "When I see a danger signal handed to me, I don't argue with it. I get out! A few days later, if everything looks all right, I can always get back in again. Thereby I have saved myself a lot of worry and money. I figure it out this way. If I were walking along a railroad track and saw an express train coming at me sixty miles an hour, I would not be damned fool enough not to get off the track and let the train go by. After it had passed, I could always get back on the track again, if I desired."

> "The problem is that financial markets generate data that is not like the data on which machine learning works well."

Because computers interpret information differently than humans, they may miss some trains coming. Recent analysis of image-recognition deep-learning algorithms, for example, reveals that tiny errors – the change of a single pixel in an image, for example – can lead algorithms to fail miserably. This raises the possibility that some market participants may create just such tiny errors in financial data on purpose as a way to change the trading behavior of other active algorithms, using a tiny perturbation of the price or other data to shift a competing trading algorithm from buy to sell or no action at all, or to increase or decrease the size of buy and sell orders.

This is an interesting turn of events. Research suggests that manipulating the prices of securities through mere trading (as opposed to fraud) is quite difficult, at least when humans oversee trading decisions [Fischel and Ross (1991), Kyle and Viswanathan (2008)]. But we know far less about how price manipulation might work in a computer-driven market, and there are reasons to believe stock manipulation is more widespread than recognized [Comerton-Forde and Putnins (2014)]. Quant trading is likely to raise important regulatory issues in the future [Korsmo (2014), Mahoney and Rauterberg (2017)]. The temptation to manipulate markets may be particularly large for some quants as they find themselves unable to generate promised returns legitimately.

4. COGNITIVE BIASES, OR, HOW TO MAKE MONEY AS A QUANT WITHOUT REALLY TRYING

One way to understand quant funds – like most hedge funds – is to realize they are not about superior investment performance. Most hedge funds charge enormous fees to deliver performance that consistently underperforms passive (and very inexpensive) index funds. Investors in hedge funds may be the dumb money in the market. They are optimistic and overconfident gamblers who think they can pick a winner notwithstanding the failures of similarly-situated investors to do so. To them, hedge funds – including quant funds – are like casinos. And like the real casino business, hedge fund investors like to frequent the shiniest and brightest. Hedge funds build their casinos accordingly: hiring lots of people with no proven ability to beat the market, but who look awfully smart. This includes data scientists.

Take a firm like AQR. Their results speak for themselves; it is not a performance powerhouse. But its founder, Cliff Asness, is a master marketer of academic research. He has an unparalleled advantage at making investors feel he is implementing for them the lessons of tried and true academic research. By hiring (co-opting?) incredibly smart academics and appealing to academic journal results, he builds a casino that attracts the gamblers who want a University of Chicago patina of academic rigor. But is he adding much value over Vanguard's cheaper index funds? It doesn't appear so. BlackRock is the world's largest asset manager, but it makes far more money on its most expensive products than the passive products that are best for customers (I think of passive index customers as the non-gamblers who come to the casino for the great food at rock-bottom prices). BlackRock has now set up a group to research artificial intelligence in investment. That seems more about marketing than anything else.

This window dressing is unlikely to generate returns for the reasons I reviewed above, but it is likely to exploit investor optimism and overconfidence and make considerable sums for the quant managers, especially those who were fortunate to generate high returns long ago on much smaller asset bases and who can, therefore, still claim they have high "average" returns. Optimism is an important cognitive bias that draws investors to overpriced active management, including quant strategies. Much psychological research shows that individuals do not base predictions upon objective evidence, e.g., the evidence that the median active manager does not beat passive indices and the evidence that the only reliable persistence in returns is that really bad active managers tend to remain really bad. In a widely cited paper in Nature Neuroscience, Sharot et al. (2011) suggest that optimism may arise because desirable information is integrated into prior beliefs more readily than undesirable information. When newly encountered information – the underperformance of your hedge fund investment in Citadel or Winton, say – is worse than expected, people largely ignore it, perhaps consoling themselves with that "average" return that places significant weight on big returns from the 1990s.

5. THE TASK IS HARDER THAN YOU MIGHT THINK

Quant strategies face a bigger problem than the limits of data science with financial data. In prior work with co-authors [Heaton et al. (2017)], confirmed by related work [Bessembinder (2018)], I find that active managers are probably doomed to underperform large passive (and inexpensive) indexes, like the S&P 500, in most years because active strategies miss the handful of stocks that drive market results. An underemphasized empirical fact is that the best performing stocks in a broad index often perform much better than the other stocks in the index, so that average index returns depend heavily on a relatively small set of winners. Quant strategies that select subsets of securities from an index are likely to underperform it.

To illustrate the idea, consider an index of five securities, four of which (though it is unknown which) will return 10% over the relevant period and one of which will return 50%. Suppose that active managers choose portfolios of one or two securities and that they equally weight each investment. There are 15 possible one or two security "portfolios." Of these 15, 10 will earn returns of 10%, because they will include only the 10% securities. Just five of the 15 portfolios will include the 50% winner, earning 30% if part of a two-security portfolio and 50% if it is the single security in a one security portfolio. The mean average return for all possible actively-managed portfolios will be 18%, while the median portfolio of all possible one- and two-stock portfolios will earn 10%. The equally-weighted index of all five securities will earn 18%. Thus, in this example, the average active-management return will be the same as the index [see Sharpe (1991)], but two-thirds of the actively-managed portfolios will underperform the index because they will omit the 50% winner. Quant strategies face a daunting task in beating the odds of missing the best performing trades. And by constant trading, they create even more positions that are likely to underperform market indexes.

6. CONCLUSION

In this article, I explore some limits of quantitative investment with a focus on the promise – or lack thereof – of techniques from deep learning and artificial intelligence more generally. In prominent applications of deep learning, the goal is typically to train a computer to do as well or better at a task – such as recognizing the content of an image – that a human usually does quite well. But financial markets present far different problems than those presented in computer vision, speech recognition, and natural language processing. Given the mostly unimpressive performance of quant funds – with a tiny handful of exceptions – it is more reasonable to view quantitative investment management as more marketing than effective trading technique. Moreover, there are empirical reasons that it is very difficult to beat large passive portfolios consistently, and those empirical facts are just as hard for quants to overcome as for other active managers. While deep learning and artificial intelligence are changing the world in many ways, they are unlikely to generate fortunes for investors, who will continue to remain best-served by inexpensive and passive index products that will be augmented by machine learning techniques to drive costs even lower.

References

Bessembinder, H. 2018, "Do stocks outperform treasury bills?" Journal of Financial Economics, forthcoming

Comerton-Forde, C., and T Putnins, 2014, "Stock price manipulation: prevalence and determinants," Review of Finance 18, 23-66

Fischel, D., and D. Ross, 1991, "Should the law prohibit 'manipulation' in financial markets?" Harvard Law Review 105, 503-553

Heaton, J. B., N. Polson, and J. H. Witte, 2017, "Why indexing works," Applied Stochastic Models in Business and Industry 33, 690-693

Livermore, J., 1940, How to trade stocks, Duell, Sloan & Pearce

Korsmo, C., 2014, "High-frequency trading: a regulatory strategy," University of Richmond Law Review, 48, 523-609

Kyle, A., and S. Viswanathan, 2008, "How to define illegal price manipulation," American Economic Review 98, 274-279

Lin, T., 2017, "The new market manipulation," Emory Law Journal, 66, 1253-1314

Mahoney, P., and G. Rauterberg, 2017, "The regulation of trading markets: a survey and evaluation," working paper

Merton, R., 1980. "On estimating the expected return on the market," Journal of Financial Economics 8, 323-361

Sharot, T., C. W. Korn, and R. J. Dolan, 2011, "How unrealistic optimism is maintained in the face of reality," Nature Neuroscience, 14, 1475-1479

Sharpe, W. F., 1991, "The arithmetic of active management," Financial Analysts Journal 47:1, 7-9

Taylor S., and J. Brown, 1988, "Illusion and well-being: a social psychological perspective on mental health," Psychological Bulletin 103, 93-210

Weinstein, N., 1980, "Unrealistic optimism about future life events," Journal of Personality and Social Psychology, 39, 806-820

SECURITY

PREVIOUS EDITIONS OF THE CAPCO JOURNAL OF FINANCIAL TRANSFORMATION ARE AVAILABLE AT **WWW.CAPCO.COM/INSTITUTE**

- Cyber security ontologies supporting cyber-collisions to produce actionable information
- Digital ID and AML/CDD/KYC utilities for financial inclusion, integrity and competition
- Digital identity: The foundation for trusted transactions in financial services
- Setting a standard path forward for KYC
- E-residency: The next evolution of digital identity
- The future of regulatory management: From static compliance reporting to dynamic interface capabilities

Cyber security ontologies supporting cyber-collisions to produce actionable information

MANUEL BENTO | Euronext Group Chief Information Security Officer, Director, Euronext Technologies
LUIS VILARES DA SILVA | Governance, Risk and Compliance Specialist, Euronext Technologies, CISSP
MARIANA SILVA | Information Security Specialist, Euronext Technologies

ABSTRACT

In this article, we bring to the attention of key players in the financial services sector the continuous cyber security events affecting the industry globally. We also consider a possible solution for mitigation of such events through the introduction of new processes and technologies. Using a computational logic-based language by machine learning processes through artificial intelligence algorithms will improve prediction of unwanted cyber events via early warning alerts. A cyber-collision system concept is described by adjoining cyber security ontologies, security analyst experience, machine learning, and information sharing to protect the financial services sector.

1. INTRODUCTION

The current cyber security paradigm identifies well-defined activities, promoting multiple and increasing layers of defense in a reactive mode. We know from experience that "intrusion prevention systems" (IPS) do fail with an excessive number of false positives, which must be minimized through the tuning of the faulty detection signatures.

The current approach has failed in preventing ransomware attacks, phishing attacks, and the old social engineering attacks that continue to cause major problems for global corporations. Case studies of such attacks are well-known to most observers of the world of business. They include the Equifax debacle, the Yahoo bombshell, the WannaCry ransomware attack, the NotPetya malware outbreak at Maersk, Fedex, and Rosneft, among others.

The information security community expects 2018 to be not too different from the recent past. If anything, the proliferation of uncontrolled systems connected to the internet through the Internet of Things (IoT) could make matters worse.

To respond to this worsening situation, firms must commit resources to attain knowledge from beyond their firewalls, so that they can predict what attacks will become likely and decide where to invest. Consequently, facing the cyber enemies outside of the "comfort zone," and being able to prevent their attacks, or, even better, being able to avoid them, is paramount to cyber security. Battle-tested machine learning processes will, with the help of specialized security professionals, improve predictive analyses. Together with proactive financial business sector involvement, they could promote a cyber-collision system to handle positive cyber attack alerts from multiple sources using a centralized cyber attack index system employed for cyber defense support.

2. CURRENT REALITY – THE KNOWN WRONGS

Currently, enterprises, organizations, and governments have major difficulties in detecting information security attacks or even reacting to them when detected, especially when they affect multiple systems in many disperse geographical locations.

Knowledge of information security attack vectors is paramount to information security analysts, as cyber attackers have the capability to learn about any online business resource and evaluate its interconnectivity with systems within the same business sector. This is especially the case among financial services firms, who are not very open about sharing data to support peers (for example on failed access attempts). Nonetheless, willingness to start sharing and even creating a common approach to cybersecurity is helping the financial services sector with, at this stage, dealing and handling cyber response to known cyber attacks (e.g., FI-ISAC and FS-ISAC: Financial Services – Information Sharing and Analysis Center, mailing lists).

Recent cyber attacks have demonstrated that only after a process of public awareness of their real impact do companies call their security analysts to report on the cyber-resilience controls in place; typically with difficult to understand dashboards. Consequently, sharing security alerts at an early stage could improve the analysis process and minimize the impact of a cyber attack.

A data breach is the most disruptive cyber attack security incident. Consequently, firms should systematically identify and sanitize key lessons from cyber events in order to advance resilience capabilities.

Per Verizon's "2017 data breach investigations report," an incident is a security event that compromises the integrity, confidentiality, or availability of an information asset. On the other hand, a breach is an incident that results in the confirmed disclosure (not just potential exposure) of data to an unauthorized party.

> "Security specialists can perform deeper investigations to improve machine learning models and help transition from a reactive approach to a proactive one, and eventually to a predictive approach."

Data breaches frequently cause major reputational damage. They create uncertainty, reduce consumer trust, and can harm the firm's competitive edge in the markets.

Technology has become a vital part of people's daily lives and is crucial for societies to grow. As a result, security awareness must be among the major investments in information security by the financial services industry.

Technology may fail, however, early warnings through information sharing could create actionable information to mitigate cyber threats. It is common nowadays to see social engineering skills (e.g., through spear phishing) being used to explore weaknesses in order to obtain access to companies' valuable information.

On the other hand, the cyber threat landscape will be mostly about new applications leveraging new business models, based on new technologies, and making use of new infrastructure models (xaaS). Further, new rapid application development technologies, associated with complex algorithms to combine disparate data sources, including the merging of internal business specific data with external "big data" analysis to expand sector/oriented workflows and processes (such as using hybrid cloud systems), will make cyber security and data protection difficult disciplines to handle within one organization. A very real example can be found in poorly designed applications being dependent on built-in OS kernel libraries with obsolete algorithms for encryption/decription that cannot always be removed due to the loss of source code. Consequently, security measures need to be integrated at a later stage. A typical example is the undocumented use of "forked" Linux kernel libraries by Java applications.

Analysis of past data breaches alone, while helping us to understand common weaknesses, is not enough to stem the tide. What is needed is predictive analysis based on massive security event datasets to identify trends, predict impacts, and propose mitigating actions. Such analysis will be based on classification mechanisms that are underpinned by cyber security ontologies and feed AI algorithms allowing the identification of cyber-collisions, such as prediction of cyber attacks through clear alerts based upon experience or knowledge.

Through identification and study of past data breaches, we will be able to establish a well-defined baseline of behavioral and system activity against which we can apply machine learning techniques. Big data analysis, helped by cyber security ontologies and based on datasets of past events, enable algorithms to be trained to learn trends and impacts and propose mitigating solutions and consequently stop cyber attacks through learning collision mechanisms.

Table 1: Biggest data breaches of the 21st Century (U.S.£ million)

2017	Equifax	143
2016	Adult Friend Finder	412.2
2015	Anthem	78.8
2014	eBay	145
	JP Morgan Chase	76
	Home Depot	56
2013	Yahoo	3000
	Target Stores	110
	Adobe	38
2012	U.S. Office of Personnel Management (OPM)	22
2011	Sony's PlayStation Network	77
	RSA Security	40
2008	Heartland Payment Systems	134
2006	TJX Companies Inc.	94

Source: Armerding (2018)

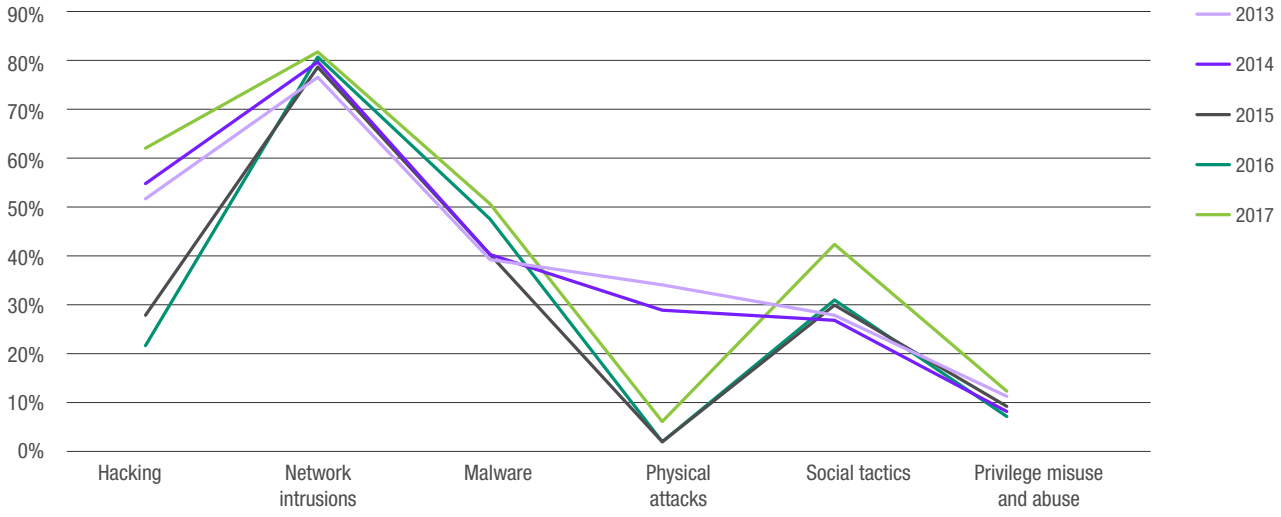
Table 1 depicts the biggest data breaches from this century.

In the financial services industry, the most prevalent types of attacks are: "denial of service attacks" (DDoS), web application attacks, and payment card skimming. Cyber attack methods are presented in Figure 1, showing clearly that the weakest link is still related to network intrusions and active hacking.

3. CHALLENGES OF DATA SCIENCE APPLIED TO CYBER SECURITY

At present, there are some challenges in applying data science to cyber security [Kolman (2014)]: data normalization, anomaly detection, high cost of errors, the data required is not public, data evaluation is difficult, semantic gap (with the difficulty to describe the information), lack of expertise, and an adversarial environment were permanently changing datasets imply a learning period for humans to adapt. Consequently, the use of data science in a cyber security context can be considered to require very specialized human skills and a large commitment of effort.

Figure 1: Verizon data breach investigations reports



As we can see from Figure 1, targeted attacks on computer networks are still the prevalent method of cyber attacks and the need for tools to support analysts to effectively hunt malicious activity within one's perimeter has increased dramatically. The existing security event information management (SIEM) systems help analysts through pre-defined schemas to identify logs or events that might be of interest within the aggregated logs. As described before, these systems are faced with some hard problems, namely the diverse schema of information sources that imply extra layers of technology – connectors – to properly incorporate information and security analyst expertise to help correlate the new source with existing log events.

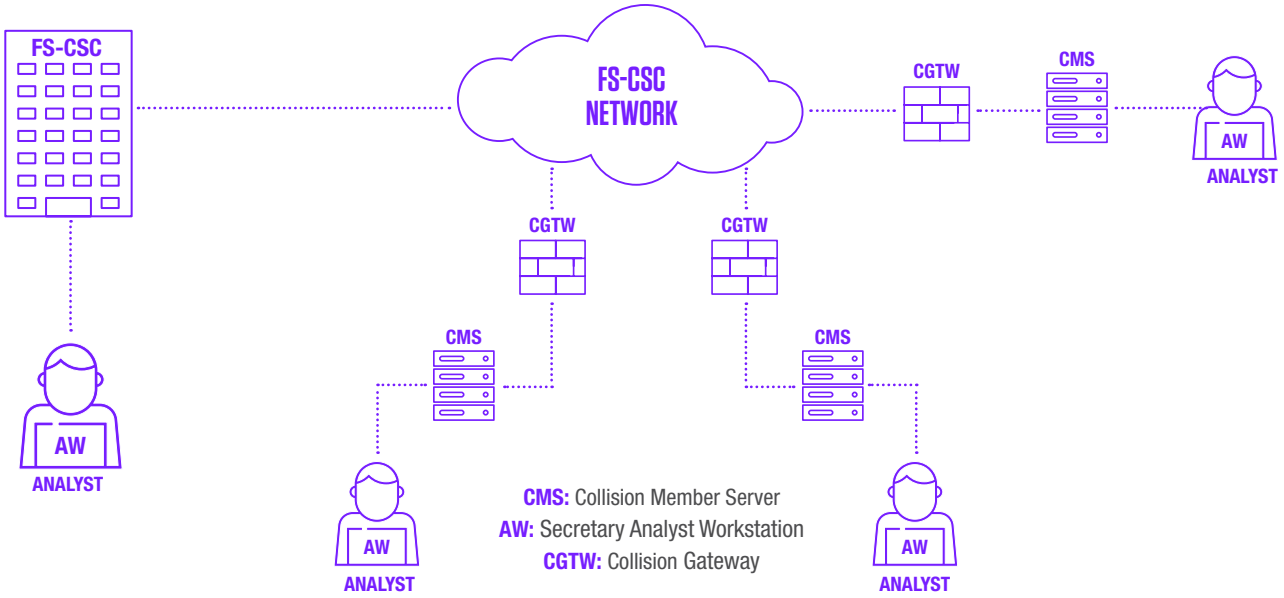
By adopting a unified way to support information integration and cyber situational awareness in cyber security systems, security analysts will be able to get better visibility on threats. As such, adopting a cyber security ontology will make available to security analysts, intrinsic properties, that with some "assumptions" (like "false positive multi location access" – for example a login at an office in London while the same login account is used within the same timeframe at HQ in Paris using mobile access to email), will overcome the workload limits, making it possible to analytically process the huge and constantly changing event datasets. The analysis process will, therefore, entail the adoption of the right learning and processing models. The process of further assigning specific metadata (i.e., required ontology labels) or attributes to identify appropriate analytical sources aiming to reduce the specialized security knowledge required for an analyst to be effective at understanding and evaluating a threat will also be incorporated.

4. MACHINE LEARNING HELPING CYBER SECURITY

Choosing ontologies as the unifying support to information integration, sharing generated cyber attack information through an event repository with heterogeneous data and knowledge schemas, can contribute to producing actionable information to feed a hypothetical cyber-collision mechanism. This proposed cyber-collision mechanism is fed by multivariate systems (systems that use incremental learning algorithms such as pattern recognition, data mining, or fuzzy logic), so attackers do not become familiar with a specific system. It is supported by security experts' cooperation to improve the models used by such systems. Machine learning systems are, therefore, the final element with highly integrated functions of high-performance analytics for predictive analysis and forecasting of cyber attacks.

Adopting the "unified cyber security ontology" (UCO) [Syed et al. (2016)], security specialists can perform deeper investigations to improve the machine learning models and help transition from a reactive approach to a proactive one, and eventually to a predictive approach. As we know from machine learning theory, experience increases task performance [p(t,e)>p(t)], which is the motivation for the

Figure 2: High level cyber-collision network alerting system



Source: Luis Vilares da Silva / Sofia Silva

cyber-collision mechanism to process the actionable information and predict cyber attacks. At the same time, if the cyber attack materializes, the model can improve situational awareness, which is of extreme importance when performing digital forensic investigations or defining a tactical cyber defense strategy.

The idea behind UCO is paramount for moving the cyber security paradigm from event correlation to an extensive cyber situational awareness in systems like the one proposed here as an example.

UCO ontology was mapped to many existing cyber security ontologies and concepts in the Linked Open Data cloud [Bizer et al. (2009)] and is an extension of the "intrusion detection system ontology." The UCO authors describe the ontology as the core for a cyber security "linked open data" (LOD) cloud as it represents the semantic version of the event exchange standard STIX, extended with other cyber security related standards, such as "common vulnerability and exposures" (CVE), "common attack pattern enumeration and classification" (CAPEC), etc.

The purpose of UCO is to serve as the core for the cyber security domain and its capacity to be extended serves the purpose of structuring event information to be shared, integrated, and reused within applications in the financial realm.

"Resource description framework" (RDF) and languages such as "ontology web language" (OWL), are used to represent entities through a set of abstract objects or concepts rather than only some strings of words. Both languages expose structures that represent information that is not only machine readable, but also machine understandable, and therefore facilitate the sharing of information from heterogeneous sources.

5. CYBER SECURITY ACTIONABLE COLLABORATION MODEL IN THE FINANCE SECTOR

Our hope is that the financial services industry can join efforts, through a consortium type of organization, to create a collaborative platform to properly predict cyber attacks through preemptive positive alerts producing actionable information enabling "early warnings" for the members of the consortium.

With such an ambition, the first set of questions arise: who will be in the consortium? Who should lead the process? What type of data should be shared? Where will the data be located? How is the communication processed in a synchronous way? Who is paying for it? Finally, which type of system are we proposing?

5.1 The consortium

To be broad but effective, the proposed consortium would include both frontline membership and consulting membership. The first type of membership in the finance sector would include trading companies and brokers, banks, credit card companies, and insurance companies. For the latter membership, international financial institutes and international law enforcement organizations are the interested parties to support the containment and undertaking further investigative assessments for proper incident handling.

The consortium should have a governing body, management structure, and governance model agreed by all stakeholders. This consortium could be described as the Financial System Cyber Security Center (FS-CSC).

5.2 Sharing data

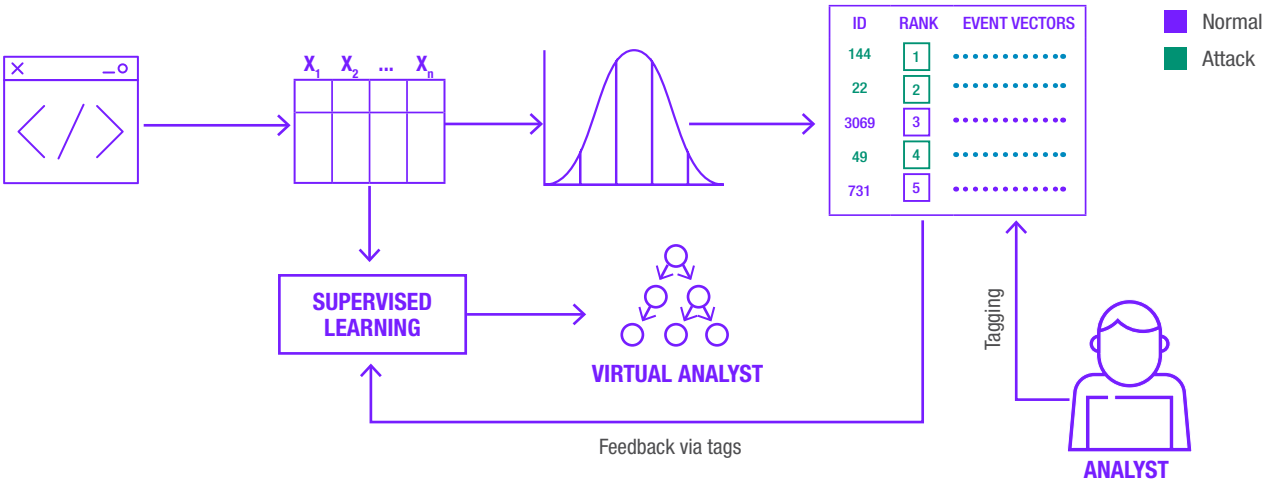
The most important element within collaborative platforms, as in any information system, is the data they process. For the proposed system and having in mind the reservations companies could have in sharing their data with competitors within the same sector, a specific and well-defined set of data would be considered. Using technology, such as intrusion prevention systems to analyze and convert scanned traffic into cyber security ontologies with features/attributes such as AttackPattern, BaseGroup, CCE, CVE, CVSSScoreType, ConfidenceType, Exploit, Malware, Origin, Attack, Attacker, Campaign, Consequence, etc... (i.e., an agreed subset of the UCO ontology), and centralize the output result in a large-scale logging system would, therefore, be a type of solution to share common cyber defense information.

5.3 Localization, network, and funding

The physical location of the event data is of major importance, with the financial services industry being heavily regulated. For a system that will share a security advantage with its members, an independent, stable, and regulated country should be chosen to host the infrastructure supporting it. Countries such as Switzerland, Luxembourg, Portugal, or Ireland would have the technology, stability, and regulations required to host an independent, advanced, data center such as this. They are also at a much lower risk of facing terrorist attacks.

For the communication to be secure, while effective, a private network should be created and configured to interconnect the security departments of the finance sector members who would also pay on a pro rata basis (whichever order of magnitude is used to classify the members).

Figure 3: Prediction system



Source: Kalyan Veeramachaneni/MIT CSAIL

5.4 System components – high level approach

From a topology perspective, the components are familiar to any InfoSec team within the industry. As depicted in Figure 2, the system has the normal business exposure to the internet, but there is a new routing mechanism for preemptive alerting of cyber attacks, which are described as Cyber-collision Gateways.

In the proposed system, each member's analyst will feed the system with tagged ontologies, allowing the learning process to improve the analysis of the traffic and, therefore, enhancing the detection of positive cyber attacks' alerts.

Ideally, the learning process will improve the analysis in a way that a virtual analyst will replace the member's analysts; a system similar to the virtual artificial intelligence analyst developed by the Computer Science and Artificial Intelligence Lab and the company PatternEx (Figure 3) that reduces false positives by factor of 5 [Connor-Simons (2016)].

Furthermore, the event database created is available to complement the alert mechanism with a search for actionable information through a hidden-hit mechanism.

The hidden-hit mechanism is a process that informs the owner of the information when his data was hit by a search and by whom. This process will allow the owner to decide if his information can be shared immediately or will trigger another process of peer communication between the searching actor and the information owner. This process is key for reporting purposes and to create security dashboards.

The key factor with this approach is that, on one hand, members do share information related to their internet traffic without sharing business information and, therefore, competitors will not take business advantage, and on the other hand, all members have access to more information with extra relevance.

6. CONCLUSION

Machine learning in cyber security will increasingly replace the current paradigm where reactive mechanisms protect our systems, but we continue to be vulnerable as the recent cyber attacks demonstrate.

Proper risk management practices will go beyond reactive controls and include proactive protection against unwanted future cyber events. The proposed approach for the financial services industry includes proper sharing of information related to internet traffic and, therefore, improve the defense perimeter. Having a common mechanism for improved alert on cyber attacks will accelerate cyber defense capabilities, which is also extremely important for advanced persistent threats. In the long run, this approach will save operational costs with a centralized virtualization of analysts.

References

Amerding, T., 2018, “The 17 biggest data breaches of the 21st century,” CSO Online, January 26, http://bit.ly/2ovBb24

Bizer, C., T. Heath, and T. Berners-Lee, 2009, “Linked data – the story so far,”
International Journal on Semantic Web and Information Systems 5:3, 1-22

Conner-Simons, A., 2016, “System predicts 85 percent of cyber attacks using input from human experts”,
MIT News, April 18, http://bit.ly/1SWb2OZ

CSO. “Biggest data breaches by year and accounts compromised”
https://goo.gl/KkNXLw – http://breachlevelindex.com

Kolman, Y., 2014, “Machine learning and big data in cyber security,”
http://bit.ly/2CMPzHE; http://bit.ly/2BVuoGs

Syed, Z., A. Padia, M. L. Mathews, T. Finin, and A. Joshi, 2016, “UCO: a unified cyber security ontology,”
Proceedings of the AAAI Workshop on Artificial Intelligence for Cyber Security

Verizon, 2017, Data breach investigations report

Verizon, 2016, Data breach investigations report

Verizon, 2015, Data breach investigations report

Verizon, 2014 Data breach investigations report

Verizon, 2013, Data breach investigations report

Digital ID and AML/CDD/KYC utilities for financial inclusion, integrity, and competition

DIRK A. ZETZSCHE | Professor of Law, ADA Chair in Financial Law (Inclusive Finance), Faculty of Law, Economics and Finance, University of Luxembourg, and Director, Center for Business and Corporate Law, Heinrich-Heine-University, Düsseldorf, Germany

DOUGLAS W. ARNER | Kerry Holdings Professor in Law, University of Hong Kong

ROSS P. BUCKLEY | King & Wood Mallesons Chair of International Financial Law, Scientia Professor, and Member, Centre for Law, Markets and Regulation, UNSW Sydney

ABSTRACT

Customer identification is key to protecting market integrity. The know your customer, anti-money laundering, and counter-terrorism financing rules all work to this end. However, these strict rules can limit access to financial services, particularly by small and medium enterprises and poorer individuals. Global interest in e-identity is growing, with multiple countries either establishing, or having already established, national e-identity systems. The potential of centralized identity databases to simplify the experience of accessing both government and financial services is clear. Efficient e-identity services also hold great potential for international financial centers. This article sets out three measures to which such centers must pay particular attention in building their e-identity systems.

1. THE CHALLENGE OF E-ID: SQUARING THE CIRCLE

The financial services sector supports economic growth and development through allocating financial resources, providing investment opportunities, and managing risks. Financial regulation seeks to promote these functions through minimizing the frequency and severity of financial shocks (financial stability), enhancing access to financial services (financial inclusion), and promoting market integrity.[1] From the standpoint of an international financial center (such as Hong Kong, Luxembourg, or London), competitiveness derives from balancing these objectives and providing the necessary infrastructure for financial markets to function well.

Verifying customer identity and carrying out "know your customer" (KYC) due diligence on acceptance of a new customer (on-boarding) and on an ongoing basis are fundamental to market integrity, as these are essential to maintaining confidence and trust in the financial system and reducing the likelihood of criminal or terrorist access to financial services. The rules for these measures are embodied in a wide range of AML/CFT/CDD requirements (anti-money laundering/countering the financing of terrorism/customer due diligence),[2] based on internationally agreed approaches.[3] In addition, CDD underpins how customer needs are understood and is essential to providing appropriate financial services, a function often summarized under the general framework of suitability.[4]

At the same time, these requirements restrict access to financial services and must, therefore, be balanced against the objectives of financial inclusion and economic growth. In particular, loss of access to the financial system restricts access to financial services for small- and medium-sized enterprises (SMEs). SMEs are central to economic growth and innovation, and reducing, or in some cases eliminating, their access to finance has important consequences for growth, innovation, and development. In addition, financial institutions, corporates, and individuals in emerging and developing markets (such as most of Asia) are often seen as "high risk" and hence subject to "de-risking," particularly by financial institutions from Western developed markets.[5] This issue has become sufficiently significant to be the focus of the G20, the Basel Committee, and FATF, among others, with one solution being to adjust standards in order to reduce the disproportionate impact on correspondent banks in emerging and developing markets (particularly Asia) and their customers.[6]

Beyond SMEs and correspondent banking, the G20 (particularly through its focus on digitally inclusive finance)[7] and the United Nations (U.N.) (in particular through the U.N.'s Sustainable Development Goals)[8] have made financial inclusion a central policy objective, on equal footing with financial stability and integrity. In this context, in addition to de-risking, AML/CFT/CDD requirements often make it difficult for underserved segments of society to access the formal financial system, particularly the poor in rural and urban areas. Financial inclusion is seen as central to supporting economic growth and reducing poverty and inequality, as it empowers individuals to improve their circumstances by using financial services, and particularly digital financial services delivered through mobile and smart phones.

Financial technology (fintech),[9] and in particular "regulatory technology" (regtech),[10] present opportunities to reconsider existing systems and to build the necessary infrastructure to balance market integrity, financial inclusion, and economic growth, while at the same time meeting commitments to international financial standards including those set

---

1 For instance, by striving to prevent the criminal or terrorist use of the financial system and limit market manipulation and misconduct; as all of this behavior impacts confidence and trust in the financial system.

2 For the E.U. rules, see the Fourth AML Directive (Directive (EU) 2015/849 of the European Parliament and of the Council of 20 May 2015 on the prevention of the use of the financial system for the purposes of money laundering or terrorist financing, OJ L 141, 5.6.2015, p. 73–117; for Hong Kong see (i) the Anti-Money Laundering and Counter-Terrorist Financing (Financial Institutions) Ordinance ("AMLO"), (ii) the Organized and Serious Crimes Ordinance ("OSCO"), (iii) the Drug Trafficking (Recovery of Proceeds) Ordinance ("DTROP"), and (iv) the United Nations (Anti-Terrorism Measures) Ordinance ("UNATMO"); for Singapore see the Monetary Authority of Singapore's various notices and guidelines on AML/CFT, available at http://bit.ly/2p5BgJX; for Australia, see Anti-Money Laundering and Counter-Terrorism Financing Act 2006 (Cth).

3 See the standards provided by the Financial Action Task Force (FATF), http://bit.ly/2f1TJAA. The Financial Action Task Force (FATF) is an inter-governmental body established in 1989 by the Ministers of its member jurisdictions. The objectives of the FATF are to set standards and promote effective implementation of legal, regulatory, and operational measures for combating money laundering, terrorist financing, and other related threats to the integrity of the international financial system. The FATF is, therefore, a "policy-making body" that works to generate the necessary political will to bring about national legislative and regulatory reforms in these areas. The FATF framework is composed of the 1) FATF Recommendations 2012, 2) international anti-money laundering and combating the financing of terrorism and proliferation (AML/CFT) standards, and 3) FATF Methodology to assess the effectiveness of AML/CFT systems 2013.

4 For the E.U., see Article 25 of Directive 2014/65/EU of the European Parliament and of the Council of 15 May 2014 on markets in financial instruments, OJ L 173, 12.6.2014, p. 349–496.

5 For instance, the Hong Kong Monetary Authority (HKMA) issued a circular on de-risking and financial inclusion on September 8, 2016 (http://bit.ly/2Im1cJv) to banks operating in Hong Kong: the HKMA observed months of media reports on the plight of some customer groups who were excluded from banking services. The HKMA warned about the dangers of screening out too many potential customers, because the resulting de-banking or financial exclusion of some customer groups could harm Hong Kong's economy and its reputation as one of the world's leading international financial centers. As a follow up, on October 11, 2017 the HKMA, Securities and Futures Commission (SFC), and Insurance Authority (IA) each relaxed their respective requirements to verify addresses in the context of AML (see Ref. B10/1C, http://bit.ly/2FByORK).

6 See "Outcomes FATF Plenary, 21-23 February 2018", FATF, http://bit.ly/2EMkwRT.

7 GPFI, 2016, "Updated G20 financial inclusion indicators focus on digital financial services," G20 Financial Inclusion Indicators, August 10, http://bit.ly/2FvXkrI

8 UNCDF "Financial Inclusion and the SDGs," United Nations Capital Development Fund, http://bit.ly/2DkTBap

9 Arner, D. W., J. Barberis, and R. P. Buckley, 2016, "The evolution of FinTech: a new post-crisis paradigm?" Georgetown Journal of International Law 47:4, 1271-1319

10 Arner, D. W., J. Barberis, and R. P. Buckley, 2017, "FinTech, RegTech and the reconceptualisation of financial regulation," Northwestern Journal of International Law and Business 37, 371-414

by the FATF, Basel Committee on Banking Supervision, Financial Stability Board (FSB), and the U.N. In this article, we examine how financial centers could make use of technology in the context of digital identity and electronic AML/KYC requirements.

This article identifies and considers three different aspects which must be addressed strategically:

- Digital ID infrastructure
- eKYC infrastructure
- Suitability infrastructure

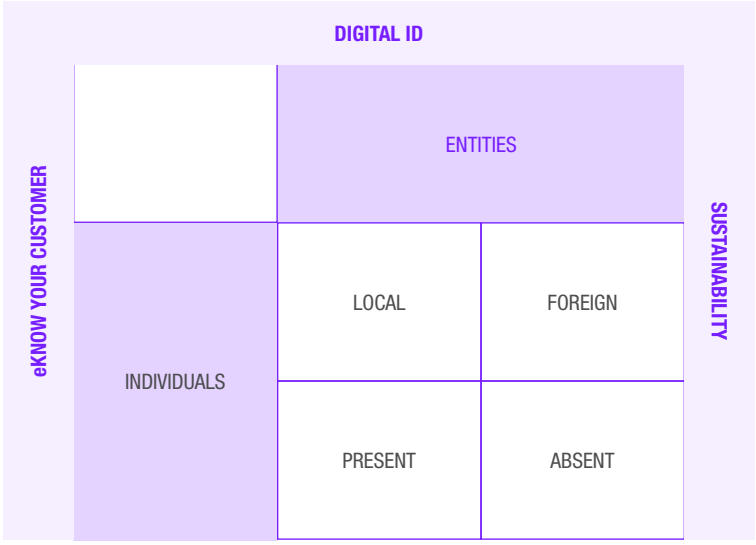
Across each of these aspects, the article considers two different contexts that must be addressed as part of the strategy: (1) individuals and (2) entities (especially companies). Within these two contexts, the strategy must also address: (1) local and (2) non-local individuals and entities, and also (1) physically present and (2) non-physically present individuals and entities. In each case, infrastructure and utilities could be built by the government, the private sector, or in some form of collaboration. Likewise, in each case, systems and utilities could be exclusive (for example, sovereign identity sources from sovereigns) or open (for example, a system of licensing for competitive providers), or something in between (for example, a licensed single provider).

This matrix lays out the central elements of a strategy for putting in place the necessary financial infrastructure to meet objectives of financial integrity, financial inclusion, and financial competitiveness, with the following sections addressing each of digital ID, eKYC utilities, and suitability in turn.

2. THE ROLE AND BENEFITS OF SECTOR-WIDE E-ID SYSTEMS

Financial institutions, fintech startups, and technology firms engaging in financial services face a key challenge in the time-consuming and complex client on-boarding process required to meet CDD regulatory requirements. CDD data are also only useful if reliable, from a trustworthy source, and up-to-date. Financial institutions must spend a lot of time and resources on refreshing and re-verifying their client information, making transactions expensive for institutions and inconvenient for clients. In addition, from the standpoint of the overall objective of protecting market integrity, data analytics from regulatory authorities and others are most effective when applied to comprehensive pools of data. As a result, not only are existing systems

Figure 1: Digital client on-boarding matrix



expensive, inefficient, and inconvenient, they are also often not overly effective in achieving the actual regulatory objective of preventing criminal or terrorist use of the financial system. In some cases, CDD requirements could even drive legitimate businesses and financial activities out of the formal financial system and into the informal financial system. A sector-wide e-ID KYC utility is a potential solution to these challenges and, unsurprisingly, the idea of a centralized KYC utility is gaining traction globally.[11]

The next section analyzes the connection between KYC utilities and digital identification systems.

2.1 E-ID on the rise

Ensuring that all steps of identification for an E-identity can be performed online and from any location is an important objective of law makers around the globe. Examples addressing each pain point in the identification network include the Aadhaar in India, probably the most up-to-date and ambitious top down eID project, the GovPass in Australia, which connects existing ID devices and turns them into an eID system, as well as the E.U. e-IDAS Regulation, which seeks to solve the issue of how to provide cross-border eID.

[11] LexisNexis, 2016, "Banks willing to collaborate on shared KYC utility," Finextra, September 28, http://bit.ly/2dyGiYp

2.1.1 Creating digital identity from scratch – the Indian Aadhaar system

India's Aadhaar system is operated by the Unique Identification Authority of India (UIDAI), and involves issuing a 12-digit randomized number to all residents of India to be used to access government services, subsidies, social benefits, banking, taxation, and insurance, among other services. Enrollment to obtain an Aadhaar number is free, and a process of biometric de-duplication seeks to ensure that only one number is generated for each individual. The Aadhaar number issued acts as a proof of identity, but is unrelated to citizenship rights, and does not identify people's caste, religion, or income. To be issued with an Aadhaar number, an individual must satisfy the UIDAI verification process, which requires various demographic and biometric information to be provided, including the individual's name, date of birth, gender, address, mobile number, email address, ten fingerprints, two iris scans, and a facial photograph.[12]

The Aadhaar system also provides for a number of methods of updating data. As the Aadhaar number can be linked to a growing number of services, this is important. Biometric data can, for example, be updated as children grow, or in the case of accidents or diseases, or, indeed, as the quality of technology improves. Such updates can be undertaken online, using a login consisting of the individual's Aadhaar number and registered mobile number, and uploading the requisite supporting identification documents, or by visiting a permanent enrollment center in person.[13]

The Aadhaar system is subject to a hotly debated constitutional challenge in the Supreme Court of India at the time of writing. It is being argued that the identity cards are a breach of privacy, and that data is being collected by third-party contractors hired by UIDAI without proper safeguards in place. It is also argued that the biometric identification techniques, fingerprinting, and iris scanning are susceptible to misuse and fraud; and there have indeed been many problems in Aadhaar's implementation.[14] In related proceedings in mid-2017, a nine-judge bench of the Supreme Court of India held that Indians have a right to privacy, however declined to rule on the constitutional validity of the system.[15]

Aspects of the Aadhaar system subject to critique include that the Aadhaar Authentication Regulations 2016 provide for transaction data to be archived for five years from the date of transaction. Aadhaar has even been described as "mass surveillance technology."[16] However, Aadhaar has also proven beneficial. For example, billions of rupees of financial benefits previously lost annually through fraud and corruption are now finding their way to the intended recipients. The Indian government claims this alone has saved an estimated U.S.$5 billion.[17]

2.1.2 Linking identity databases – the Australian GovPass project

Australia lacks any form of national identity card, in part because earlier attempts to introduce such an initiative proved to be highly problematic politically. Identity in Australia today is generally established by reference to documents ranging from passports to drivers' licenses, and by numbers issued for tax purposes or access to Medicare. In response, the Australian Government Digital Transformation Agency (DTA) has produced the Trusted Digital Identity Framework (TDIF), a draft of which was released for public feedback in November 2017, and which is under development at the time of writing. The DTA is also undertaking a project, currently in its beta stage, to produce a digital ID for individuals to easily and securely prove their identity to government services online – the Govpass. Essentially, the technology involves using an "exchange" as a mediator between government departments and a verifier vouching for a user's identity. Once a user receives a "tick of approval" from an accredited verifier, they will be able to access available government online services. In 2018, the DTA is testing TDIF and Govpass frameworks.[18]

In October 2017, the Council of Australian Governments (COAG) reached an agreement that a national scheme should be introduced allowing for biometric identification and matching "to promote the sharing and matching of identity information to prevent identity crime … while maintaining robust privacy and security safeguards."[19] The Identity-Matching Services Bill 2018 (Cth) was introduced to the Australian parliament

[12] About Aadhaar, Unique Identification Authority of India, http://bit.ly/2HsyzJd
[13] Aadhaar data update, Unique Identification Authority of India, http://bit.ly/2xoDhG4
[14] Live Law News Network India, 2018, "SC constitution bench to begin final hearing on validity of Aadhaar cards tomorrow," January 16, http://bit.ly/2p866kw
[15] Puttaswamy (Retd.) & Anor v Union of India & Ors (Civil) No 494 of 2012.
[16] Abraham, S., R. S. Sharma, and B. J. Panda, 2017, "Is Aadhaar a breach of privacy?" The Hindu, March 31, http://bit.ly/2BpbVyx
[17] The Economist, 2016, "Indian business prepares to tap into Aadhaar, a state-owned fingerprint-identification system," December 24, http://econ.st/2FyB0hb
[18] Govpass, Australian Government Digital Transformation Agency, http://bit.ly/2Go0z1C
[19] COAG, 2017, "Intergovernmental agreement on identity matching services," Council of Australian Governments, October 5, http://bit.ly/2p5g5YO.

in February 2018. If passed, the bill will authorize the Department of Home Affairs to facilitate communication between agencies with the creation of five identity-matching services.[20] The bill also establishes the NDLFRS (National Driver Licence Facial Recognition Solution) and an interoperability hub to act as a "router," matching requests with facial image databases operated by the various services above.[21]

2.1.3 Towards cross-border digital identity: The European e-IDAS regulation

In contrast to Australia, Canada, and the U.S.,[22] identity cards with a chip embedded and common security features including the E.U.-wide use of biometrics are widely spread and used in E.U./E.E.A. member states and shared among member states' authorities. In most countries, ID cards have substituted passports and driver licenses for ID purposes.

Initially, this was also true for the U.K., where resistance against a pan-European standardized ID card was traditionally fierce. In fact, the U.K. Presidency of the E.U. council advanced E.U.-wide ID card standards, data retention, and intelligence sharing to fight terrorism in 2005, following the bomb attacks on the London subway system on 7 July 2005.[23] Following the epeal of the British Identity Cards Act by the Identity Documents Act 2010,[24] the British ID cards introduced only in 2006 were canceled. Since then, foreign nationals from outside the E.U. have been required to have an identity card, thereby turning the U.K. into something of a pre-ID state similar to that of Australia, Canada, and the U.S.

At the same time, a focus of European policy is on ensuring cross-border business transactions. European policy actions since the mid-1990s have been focused on trying to ensure that digital signatures and related declarations of will are recognized **across borders**. Since then, member states had to ensure that advanced electronic signatures based on a qualified certificate and created by a secure-signature-creation device were deemed valid signatures under the laws of each member state, in the same manner as a handwritten signature, regardless of its electronic form; in particular, digital signatures were admitted as evidence in legal proceedings.[25] However, while good in theory, in practice the e-signature received little recognition. Achieving the e-signature certificate was burdensome, few recipients had the technology to identify the certificate, and after more than a decade the technology underlying the directive was outdated. Further, the directive did not

[20] These include the FIS (face identification service), FRAUS (facial recognition analysis utility service), FVS (face verification service), IDSS (identity data sharing service), and OPOLS (one person one license service).
[21] Identity-Matching Services Bill 2018 (Cth) s 7(3).
[22] See on the U.S., Quarmby, B., 2003, "The case for national identification cards," Duke Law and Technology Review 1, 1-10.
[23] See eGovernment news – 14 July, 2005 – E.U. and Europe-wide – Identification & Authentication/Justice and Home Affairs, http://bit.ly/2FDfWSi
[24] See http://bit.ly/2FJybsP
[25] See Article 5 of the Directive 1999/93/EC of the European Parliament and of the Council of 13 December 1999 on a Community framework for electronic signatures, OJ L 13/12 of 19 January 2000.

deal with authentication and trust services, two pillars of eminent importance in today's online markets.

These issues have become particularly evident in cross-border transactions and were seen as barriers to completing the European internal market: national online trade (42%) as well as U.S.-based online services (54%) relying on enterprise-made identification systems dominate the European online economy, where E.U. cross-border online services represented a meager 4% of online sales.[26] The European regulators adopted the **eIDAS regulation (eIDASR)**[27] in 2014 with a view to reducing the costs of changing one's online relationship, be it in commerce or financial services, and enhancing competition.

The eIDASR shall provide "a predictable regulatory environment to enable secure and seamless electronic interactions between businesses, citizens and public authorities."[28] The underlying rationale is that legal certainty on eID services will assist businesses and citizens to use digital interactions as their natural form of interaction. Rather than introducing a pan-European ID card system, which would double the efforts for member states, the eIDASR seeks to ensure that people and businesses can use their own national eIDs to access public services in other E.U. countries where eIDs are available to create an European internal market for eTrust Services by ensuring that eIDs work across borders, and have the same legal status as traditional paper based processes.[29] Use cases include the submission of tax declarations, enrolling in a foreign university, remotely opening a bank account, setting up a business in another member state, authenticating internet payments, and bidding for online calls for tender.

Prior to the adoption of the eIDASR, many different national standards of eIDs were developed within the E.U. member states, independent from coordinated E.U. policy. Rather than harmonizing those standards, the eIDASR focuses on technical interoperability of all existing eID standards. By mandating the liability of member states as well as the eID provider for meeting certain identification obligations (including that the person identification data uniquely represents the person to which it is attributed and that online authentication is available),[30] the eIDASR creates trust in the eIDASR-based cross-border identification.

The eIDASR is a role model among the eID projects since it provides, in principle, an open standard not limited to E.U. jurisdictions. Every national ID system that is willing to connect to the eIDAS system could do so. Connecting to the eIDASR does not require a reform of national eID standards. Rather, by defining nodes (so-called eIDAS connectors) that provide the cross-border links between other countries' systems and one own's system any country could link to the eIDAS identification system in the E.U./E.E.A.

While adopted in 2014, the implementation of the eIDASR took some time, with public eID systems taking the lead. However, in November 2017 the first private sector-run national eID scheme was notified to the European Commission by Italy, connecting all eIDs created by private enterprise to the European eID network. This enables Italian citizens and businesses to use their Italian eID credentials to access public services in other member states.[31]

2.1.4 Sector neutrality

These ID systems are, from a sectorial perspective, neutral instruments. Financial services were not the center of attention, nor was their necessity considered, when agreeing on standards and developing technologies. For instance, the European e-IDASR tackles the issue of ensuring that a person claiming an identity is the person they say they are, with a particular focus on cross-border identification. No further information is forwarded and certified than that necessary for identification. Examples of information that is not forwarded include whether the person is a politically exposed person under money laundering legislation, or whether the person is a sophisticated or non-sophisticated investor. Further, the specific focus on identification may ignore the needs of businesses who are interested in immediate identification and authorization to link their clients to on-boarding systems. In some markets, this has led to additional (partially digital) solutions for online businesses, such as the online identification process whereby German, Luxembourg, and Swiss financial regulators allow an agent to check the identity of retail clients connected

[26] See Government of the Grand Duchy of Luxembourg, Countdown to eIDAS, http://bit.ly/2FOlUmU
[27] Regulation (EU) N°910/2014 on electronic identification and trust services for electronic transactions in the internal market (eIDAS Regulation), OJ 257/73 of 28 August, 2014
[28] European Commission, http://bit.ly/2p9FH5P
[29] European Commission, http://bit.ly/2p9FH5P
[30] See Article 11 of the eIDAS Regulation.
[31] European Commission, First private sector eID scheme pre-notified by Italy under eIDAS, 7 December 2017, http://bit.ly/2DmVQtV, online http://bit.ly/2DmVQtV.

to them via a screen camera,[32] while corporate clients must have a Legal Entity Identifier (LEI) when entering into financial services contracts.[33]

2.2 Synergies and scale economies of sector-wide e-ID utility

While compromises between digital and physical services are necessary for progress, they do not represent the "end of history." Identification is important. In theory, it is the basis for any other digital-only activity. In practice, physical identification often substitutes for e-ID where e-ID is too complex, and once physical identification occurs, intermediary-made substitutes for identification such as PIN/TAN codes distributed to smart phones, and fingerprint and iris scans reduce the importance of an efficient e-ID. Hence, e-ID can be bypassed at little cost.

More importantly, focusing on only identification, and ignoring sector-specific needs and use cases, misses many of the opportunities an e-ID system could provide. In an ideal digital services world, not only would identification proceed smoothly, but every step necessary for client-onboarding and back-up checks would be done simultaneously, and only **one time per client for all kind of services and intermediaries**. Only if this is achieved will financial intermediaries benefit from the full potential of a sector-wide e-ID system.

For instance, additional information to be embedded for financial services providers into, let's say, the LEI or a new smart ID card, could include information on links to exposed political persons (1 = yes, 0 = no, plus country identifier) and the range of financial services deemed suitable for the entity (10 = all, 9 = complex derivatives to 0 = state bonds only). This data would be machine readable and also determine which client relationships will be subject to additional checks. Once established, the receiving financial institution would tap into the KYC utility only to check whether new information is available; and these types of checks can also be fully automated, rendering manual intervention unnecessary.

The information embedded in the transaction code will not always be collected by the same entity. For instance, the payment service provider that accepts the client's money for the first time within a jurisdiction (let's say the E.U. or Hong Kong) may review the AML questions, while the first investment firm selling the client investment products may add the information on suitability. As accountability is vital, records of who has added which information and when are essential.

2.3 Responsibility

One issue facing the one-stop-shop concept for e-ID, including CDD and other financial services information, is who must take responsibility for compliance. While financial institutions may rely upon an intermediary to perform any part of the CDD measures, the ultimate responsibility for ensuring CDD requirements are met remains with the financial institution.[34] Even if a financial institution relies on CDDs performed by other intermediaries, the **respective rules of each jurisdiction are burdensome**. For instance, under Hong Kong law, the financial institution must obtain written confirmation from the intermediary that it agrees to perform the role and that it will provide, upon request and without delay, a copy of any document or record obtained in the course of carrying out the CDD measures on behalf of the financial institution. The financial institution must also ensure that the intermediary will comply with the AMLO record-keeping requirements, and if requested by the financial institution within a period of six years following the end of any business relationship with a customer, provide a copy of any document, or a record of any data or information, obtained by the intermediary in the course of carrying out CDD as soon as reasonably practicable after receiving the request. In the same vein, Article 27 of the European AML Directive requires that when financial institutions rely upon information from a third party for meeting any part of the CDD requirements, the financial institution take "adequate steps to ensure that the third party provides, immediately, upon request, relevant copies of identification and verification data and other relevant documentation on the identity of the customer or the beneficial owner."

However, the restrictions are somewhat loosened as one AML CDD can serve many banks, if a respective amount enters a bank account and only circulates within a regulated banking system where all participants are subject to the same AML rules. For example, money enters the E.U. banking system from a bank account in the Cayman Islands. The first E.U. bank needs to apply full CDD. In the absence of new information, banks that receive payments from that first E.U. bank can categorize those transactions as "low risk," i.e., they

[32] The technique was first introduced in 2015 and 2016 and clarified in later regulatory releases. See for Germany Bundesanstalt für Finanzdienstleistungsaufsicht, Circular 3/2017 (GW) – video identification procedures, Ref. GW 1-GW 2002-2009/0002, Date: 10 April 2017, online http://bit.ly/2x17fAS; for Luxembourg, CSSF, FAQ on AML/CTF and IT requirements for specific customer on-boarding/KYC methods, Version of 8 March 2018, http://bit.ly/2GlsP4M; for Switzerland see FINMA circular No. 2016/7 on video and online identification, 3 March 2016.

[33] See Article 26 of the Regulation (EU) No 600/2014 of the European Parliament and of the Council of May 15, 2014 on markets in financial instruments (MiFIR).

[34] See, for instance, Article 25 (1) of the European 4th AML Directive (supra note 2).

can in principle trust that the CDD applied by the first E.U. bank led to accurate results, and that the money is "clean."[35] The same logic could be utilized for a sector-wide e-ID plus system (or KYC utility). Note that this logic only works in closed systems, from which money cannot leak in or out.

3. TOWARDS "E-ID PLUS": SETTING UP A KYC UTILITY

The costs savings expected from an e-ID plus utility are greatest when most financial institutions participate. This statement is unlimited, in geographic terms. From an efficiency perspective, therefore, the optimum would be one global KYC utility with a full, up-to-date register of all clients within the regulated banking system.

3.1 The complexity issue

However, those who seek too much will achieve nothing. Any KYC utility project must necessarily start small. This is because hundreds of small questions must be answered to build it. Some sample questions illustrate what may be required to build a well-functioning KYC utility:

1. Which technological platform? A centralized ledger or a distributed ledger?[36] Ensuring simultaneous access is the strongest argument in favor of using distributed ledgers, while data privacy and governance concerns may tip the tide in the direction of concentrated ledgers.

2. Who shall participate and how? Answers will depend on the sophistication of technology required for participation, access to hyper-fast data streams, and reliability when performing CDD.

3. What type of information will be shared? Options include the synthesized result (i.e., "client is clean: yes/no") or variants of additional information on the client. The answer to the responsibility question raised above (II.3.) will be influential in determining how much information will be shared.

4. How often will the information be updated, and by whom? Options range from centralized data maintenance to member-based maintenance. The answer will depend on Question 2. The more reliable the members, the more acceptable is member-based data maintenance.

5. How will liability be shared if, and when, things go wrong? Options range from locating liability in one entity to joint liability. Again, this answer depends on that to Question 2. The more reliable and financially stable the members, the more acceptable is joint liability. If only the largest institutions underwrite the KYC utility, the argument for joint liability lies in incentivizing all members to invest in the maintenance and further development of the utility (similarly to how stock exchange participants together, by virtue of joint liability, are incentivized to maintain the AAA-rating of the central counterparty since its AAA rating reduces the costs of all trading partners).

[35] See Joint Committee of the European Supervisory Authorities, Joint Guidelines under Articles 17 and 18(4) of Directive (EU) 2015/849 on simplified and enhanced customer due diligence and the factors credit and financial institutions should consider when assessing the money laundering and terrorist financing risk associated with individual business relationships and occasional transactions – The Risk Factors Guidelines, JC 2017/37 of 26 June 2017, at Title III, Ch. 1 (Sectoral guidelines for correspondent banks), No. 81, 83.

[36] See on distributed ledgers Zetzsche, D. A., R. P. Buckley, and D. W. Arner, 2017, "The distributed liability of distributed ledgers: legal risks of blockchain," University of Illinois Law Review, 2017-2018, Forthcoming; Available at SSRN: https://ssrn.com/abstract=3018214 or http://dx.doi.org/10.2139/ssrn.3018214

6. Which standards will be used for data sharing? Options include an open standard or a standard designed specifically for participants.

3.2 Efficiency curve

Small improvements in this field can yield significant benefits. For instance, assume that five members each invest two staff hours in the same client. If a KYC utility is (in addition to the one-time technical set-up costs) able to reduce the needed efforts to two hours invested by only one entity, the overall cost savings approach 80%. Compare this with ten members: putting the cost of the technology aside, the costs saving would be 90%, but only 10% greater than those of the utility with five members. Those additional 10% will be partially offset by the additional costs of coordinating the additional five members. However, the calculated savings materialize only when participating institutions serve the same client. If we assume that all participants serve the same number of regional distribution of clients, the likelihood that this will be the case increases with the number of participants in the KYC utility. Under the conditions set out, the larger the utility in terms of members, the greater the likelihood of efficiency gains. Nevertheless, agreeing on governance features and standards is far easier with fewer rather than more members.

Thus, financial centers should aim to start small with a KYC utility, and plan for it to grow over time.

3.3 Reducing complexity

Legal factors may influence complexity. For instance, regulated entities are easier to include than non-regulated ones, individuals raise different questions than legal entities, and foreign financial institutions are more difficult to integrate than domestic ones, in particular foreign institutions from jurisdictions with different legal systems.

A sector-wide e-ID solution could first aim at digital identification of domestic licensed financial intermediaries, then include locally incorporated companies (relying on LEIs) and finally be utilized for non-face-to-face on-boarding of individuals. Internationalization, including foreign institutions, is perhaps the final step to be tackled.

4. GOVERNANCE OF E-ID SYSTEMS

Governance is key. This is true for any company, and particularly true for a KYC utility. Because knowledge means power, concentrating knowledge concentrates power. Take for instance, the largest global distribution center for investment funds, with its funds offered in more than 70 countries around the world.[37] A sector-wide AML/KYC tool that truly covers all client relationships will provide enormous synergies, but also pose new risks for clients globally.

How these risks could be addressed requires careful thinking that takes into account legal factors (such as property rights, liability, competition and antitrust concerns, and also applicable data privacy rules, such as GDPR)[38] together with non-legal factors (such as the technology used – with blockchain a natural candidate),[39] the cyber-security risks incurred, and the need to build a networked infrastructure to which hundreds, if not thousands, of entities can be linked.

From a governance perspective, the following legal questions are of particular importance:

1. Should the KYC utility be a public or private enterprise? A public enterprise offers public risk control, but probably also public tardiness, while a private enterprise may provide less of a long-term sustainability solution.

2. Should the KYC utility be a for-profit entity or an association acting on behalf of its members? The answer will depend in part on how the utility is to be financed. User fees could provide for ongoing maintenance costs, but up-front costs will be substantial. Given the utility will function as a monopoly, a for-profit entity with closed membership will prompt antitrust concerns.

3. Who should run the day-to-day business of the utility? This may include decisions on technical standards and the further development of the utility in light of changing technical and legal preconditions.

4. Shall the users or members have participation rights, and if so, how? Those with the greatest interest in the functioning of the utility may well have the greatest say. Voting rights could be assigned by (1) how

[37] See ALFI, 2018, "Global fund distribution," http://bit.ly/2pagnwC

[38] General Data Protection Regulation (GDPR), Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC), O.J. L119/1 of 4 May 2016.

[39] See on Blockchain Zetzsche (2017), supra 36.

often a member updates KYC data (if any), (2) how often a member requests KYC data, (3) a mix of the two, or (4) how much liability for the utility a member bears.

5. Who decides upon membership applications? The decision could be granted to an expert committee, the KYC utility's board (if any), the membership assembly, or a state institution (such as the financial regulator). Given that the reliability of members affects the utility, and the utility's financial capacity influences all members' costs, a multi-step approach requiring the recommendation of an expert committee before membership being approved could be a good process.

While research into how to set up a KYC utility is in its infancy, we believe that such utilities, to a large extent, pose **similar questions to stock exchanges in the 19th century**, since both are set up to reduce the costs of information asymmetries, and both entail a certain degree of influence on market participants. The different rules for stock exchanges around the world suggest that a one-size-fits-all answer to the questions above is impossible, and that every jurisdiction interested in KYC utilities must answer these questions for itself in light of its traditions, legal structure, and the risks its members are willing to take on.

5. THREE STEPS TOWARDS A SECTOR-WIDE E-ID UTILITY

While no single solution will address all the various issues identified, financial centers can nonetheless develop a strategic approach based on a clear understanding of existing regulation and infrastructure, international requirements, and the potential of solutions from both a technological and regulatory standpoint to address objectives, problems, and challenges. Any such strategic approach must be structured according to the needs and individual characteristics of the center. Three steps are of particular importance.

First, where a financial center is implementing **new e-identity solutions** (such as the new smart Hong Kong ID card for individual digital identification purposes, or the LEI required under MiFID for financial transactions), it is advisable to think further ahead and link such identity devices to AML/KYC checks, by ensuring that complementary technology is implemented on the side of users and that sufficient data points exist in the storage devices (in the case of LEI, this could mean that the number for the LEI is larger to include AML/KYC scores).

Second, 100% e-ID coverage is neither feasible nor likely in the short term, and aiming at 100% coverage from the beginning will either increase the risk of disruption, or delay any synergies from sector-wide e-ID systems for the foreseeable future. Thus, **complexity should** determine which steps should be taken and in which order. For instance, complexity tends to be higher on a cross-border basis and lesser on a domestic basis, and it is more difficult to include non-regulated entities than regulated ones that regularly use financial services. A sector-wide e-ID solution could first aim at digital identification of licensed financial intermediaries, then include locally incorporated companies (relying on LEIs) and finally be utilized for non-face-to-face on-boarding of individuals.

Third, from the beginning, putting a great deal of attention into the **governance** of the sector-wide e-ID tool is of utmost importance. Knowledge is power, and where there is a lot of knowledge, there is a lot of power. In particular, in global financial centers a sector-wide AML/KYC tool that covers all client relationships will provide enormous synergies, but also pose new risks. How these risks might best be addressed **requires careful thinking** that takes into account **legal factors** (such as property rights, liability, competition, and antitrust concerns, but also applicable data privacy rules, such as GDPR) and also **non-legal factors** such as the technology used (with blockchain being a natural candidate), the cyber-security risks incurred, and the need to ensure further technological evolution of a networked infrastructure to which thousands of entities may need to be linked.

Digital identity: The foundation for trusted transactions in financial services

KAELYN LOWMASTER | Principal Analyst, One World Identity
NEIL HUGHES | Vice President and Editor-in-Chief, One World Identity
BENJAMIN JESSEL | Fintech Advisor to Capco

ABSTRACT

Navigating the digital economy has become a central component of daily life – for consumers and service providers alike. The sweeping transition from the physical to digital world has fundamentally altered the ways in which organizations transact with each other, with customers, and with regulators. This has given rise to an array of new economic possibilities, increased disintermediation, and improved user experience.

Digital technologies allow people and entities to complete high-value transactions, often without ever physically interacting. With that convenience, however, comes a key question – in a digital world, how do you know that someone is who they say they are? And beyond that initial verification, how can organizations make the critical decision to trust their counterparty? Establishing a degree of assurance that someone actually is who they are expected to be, and will do what they are expected to do is an analog problem thrown into sharp relief by the volume, velocity, and complexity of modern transactions.

The digital economy has a digital identity problem. Even though identity processes are at the core of nearly every transaction individuals and institutions undertake, most identity use cases still rely on legacy paper-based credentials. These are expensive, unsecure, and will become increasingly difficult to keep compliant as new data protection regimes emerge. For financial services institutions in particular, making effective use of digital identities is both a persistent challenge and a unique opportunity. A number of innovative models have begun to emerge to more efficiently create, verify, authenticate, and federate identity information. These distinct digital identity processes lay the foundation for enduring trust with consumers, reliable compliance with shifting regulatory regimes, and continued relevance in our brave new connected economy. Moreover, as established organizations in a highly-regulated, identity-centric industry, financial institutions are uniquely positioned to drive the development of a cross-sector identity ecosystem to address both current and future digital identity challenges.

1. INTRODUCTION

Since the mass adoption of the online channel in the 1990s, the financial transactions performed by individuals and companies have exploded in value, volume, and complexity. The internet has removed many of the barriers that used to exist in exchanging goods and services, as well as in moving money between individuals and companies.

The connected economy has not only transformed traditional financial and commercial transactions, but has also facilitated the rise of new transaction types. Peer-to-peer lending and credit products, mobile payments, and automated personal financial management providers, among other innovations, do not require legacy financial intermediaries. This financial technology (fintech) revolution has been a boon to consumers, who have benefited from increased access to financial services, lower transaction costs, and far less friction than they would have encountered in visiting a physical bank branch or even calling a customer service hotline.

But, even with this wave of fintech innovation, the identity problem remains. That is, how can financial institutions assert with confidence that an individual or organization they are transacting with is who they claim to be?

That enduring question is at the foundation of trusted transactions in financial services. Fintechs and legacy institutions alike are now navigating the uneasy intersection between providing a fully digital user experience and still relying on traditional physical channels to verify and authenticate counterparties. Moving forward, effective digital identity processes will become a necessary component of a connected financial services infrastructure.

In this article, we will first explore what a digital identity is and why it is central to modern financial transactions. We will then examine the particular identity-related challenges that organizations and individuals face as they look to conduct trusted financial transactions, and highlight some innovations in the digital identity space that aim to solve these challenges. Finally, we will look ahead at the unique opportunities financial institutions may have to drive cross-sector adoption of digital identity ecosystems and facilitate future development in the space.

2. THE NEED FOR DIGITAL IDENTITY

Currently more than 60% of American consumers bank primarily online,[1] with estimates indicating that over 70% of internet users in the U.S. will use digital banking by the end of 2018.[2] In a world where the majority of financial transactions are moving to a digital channel, digital identity will have enormous consequences. Digital identity is a multi-dimensional challenge that underpins not only financial transactions, but also access to a wide array of online services.

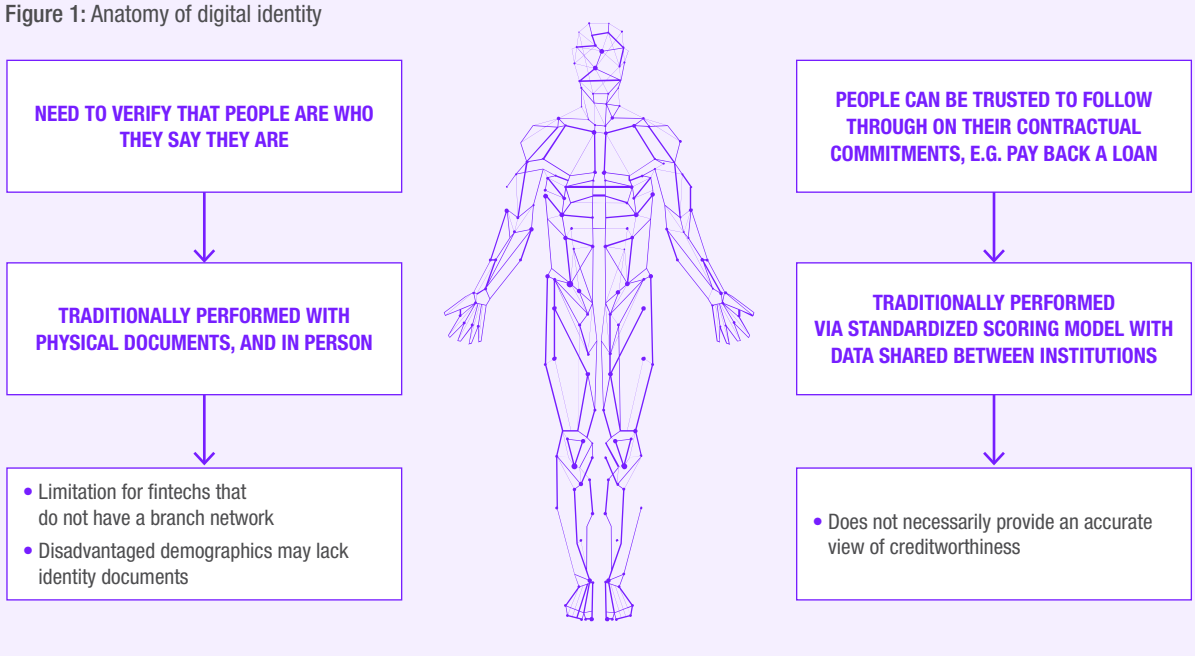
The digital identity challenge in financial transactions is far-reaching, but we will examine it here in the context of two broad, interrelated issues – verification and trust.

First, the ability to confirm that a counterparty really is who they claim to be is a primary component of transaction legitimacy. From a regulatory perspective, compliance with existing "know your customer" (KYC) and "anti-money laundering" (AML) statutes requires accurate identity verification. Even in today's digital economy, however, a consumer looking to open a checking account or apply for a mortgage often must provide physical documents in order to verify their identity and create a record with their financial institution. These legacy verification procedures are often expensive for service providers, inconvenient for users, and time-consuming for all involved. This is particularly true for markets in which traditional identity documents and credentials are hard to come by. Verification is also, in many cases, repetitive and localized to a particular service. That is, a customer must often undergo repeated checks of the same information, often requiring in-person appearances with physical documents to access different services. By the same token, financial services providers are left with the burden of secure storage or destruction of "personally identifiable information" (PII), presenting additional potential security and compliance issues.

This enduring reliance on physical identity verification also presents an especially targeted challenge for emerging fintechs. These organizations typically do not have a physical branch network and are aiming to deliver a direct-to-consumer online- or mobile-only experience, highlighting the urgent need for effective identity verification in digital channels.

[1] Statista, 2018, "Share of American population primarily using digital banking from 2014 to 2016," http://bit.ly/2BYkfXu; HM Treasury's 2015 Budget Report, March 18, 53 (Section 1.204), 98 (Section 2.272)
[2] Statista, 2018, "Penetration of digital banking among internet users in the United States from 2013 to 2018," http://bit.ly/2oGrjBY

Figure 1: Anatomy of digital identity



Second, financial institutions rely upon effective identity processes to establish counterparty trust. Confirming trustworthiness establishes a level of confidence that a customer or partner organization will actually carry out their obligations as mutually agreed in a given transaction. When counterparty trust is low or difficult to confirm, some form of recourse (either legal or through holding collateral) can provide protection in the case that one of the parties does not follow through on their obligations. Either way, an accurate evaluation of counterparty trust is contingent upon an accurate understanding of counterparty identity.

Traditional financial institutions have tended to approach trust assessment using a very limited set of identity data.[3] Evaluations of creditworthiness typically rely on a decades-old credit scoring model (like FICO) to make determinations on whether to enter into a transaction that incurs a level of risk on the bank's behalf (Figure 1). Legacy scoring models are blunt instruments, however, that exclude millions of people worldwide, especially younger consumers or those in developing markets, who may not have the credit history or physical identity documents to be "scorable" by traditional financial institutions.

Moreover, the centrality of these traditional financial institutions is being eroded in the digital economy. Increasingly, counterparties in a digital financial transaction are not banks, but rather another individual or entity. This is especially apparent in the sharing economy, where individuals are starting to monetize the excess capacity of their assets, including property (as with Airbnb or HomeAway), ride sharing (like Uber or Lyft), or even peer-to-peer lending networks (like LendingClub or Prosper). The success of these platforms is rooted firmly in trust established by a firm confidence in counterparty identity. In order for an Uber transaction to take place successfully, for example, a rider must have confidence that the person picking them up is, in fact, the correct driver, that the driver is properly licensed and insured, that the rider has entered valid payment information, and that neither the driver nor Uber itself will improperly exploit the wide array of identifying information the rider has shared (including payment information, mobile number, or location data). Each of these is a distinct identity use case that relies entirely upon the efficient, secure, and entirely digital processing of identities.

[3] For additional information on identity data and trust assessment, see OWI, 2017, "Bad credit? No credit? Big identity problem: the definitive primer on identity data in credit scoring," One World Identity, July 25, http://bit.ly/2CNV8Wf

3. ANATOMY OF A DIGITAL IDENTITY

Digital identities, then, are at the core of nearly every interaction between individuals, companies, and even devices, as the "internet of things" (IoT) continues to expand. Users rely on a variety of identities depending on the transaction at hand. The digital identity used for a Facebook profile, for example, relies on substantially different attributes, review procedures, and access protocols than the digital identity a bank uses to establish a new customer account. In order to understand how financial institutions can best apply emerging technologies to this complex problem set, it is worth dissecting the anatomy of digital identity processes in more detail.

The problem of digital identity involves multiple distinct processes that broadly encompass what attributes can be used to identify an individual, how to prove them over time, when to share them, and what a person can do with them. Given that digital identity is a broad topic, we need to define it with an additional level of granularity via the basic framework shown below, which provides five core digital identity use cases, along with the challenges and priorities inherent in each.

CREATION	VERIFICATION	AUTHENTIFICATION	AUTHORIZATION	FEDERATION
An authoritative process demarcating a particular attribute or set of attributes of an individual, entity, or object (e.g. contract, website, property, bank account), such that the attribute(s) can be used in future transactions to prove existence and uniqueness	The process of confirming at least one attribute of an individual or entity, either through self-attestation or third-party confirmation	The process of determining that one is transacting with the same entity iteratively over time	The process of determining what rights or privileges an individual or entity should be granted	The process of conveying an individual's or entity's verification, authentication, or authorization information to another party
Who are you?	How do we prove who you are?	How do we know it's still you?	What do you get once we know it's you?	How can we tell other people it's you?

3.1 Creation

Identity creation, the process of establishing trusted credentials that can be used in future transactions, is the first step in the digital identity lifecycle. Creation is an authoritative process demarcating a particular attribute or set of attributes of an individual, entity, or thing, such that the attributes can be used in future transactions to demonstrate the existence and uniqueness of that individual, entity, or thing.

For most individuals in the world, identity creation takes place in the form of government birth registration. For example, in the U.S., birth registration catalogues several attributes – name, gender, date and location of birth, and citizenship – that are fundamental to identity-related transactions throughout a person's lifetime. Governments may also mandate other identity creation processes, such as the creation of a national identification number to access benefits or pay taxes, or a motor vehicle licensing authority that can create attributes such as the type of driver's license or license restrictions. Often these same hard-copy government credentials, issued as part of basic civil registration, are required to create new records or apply for accounts with financial institutions.

Agreeing on an schema of attributes to collect for organization-specific identity creation processes can be challenging, especially when standardizing transactions internally across different departments or regions. For example, the due diligence requirements of a financial institution in Thailand are very different from those in the U.S. Similarly, it is very common for international banks with Swiss entities to interpret the attributes of an owner of a bank account in a very different way than would a U.S. division.

Identity creation involves collecting information on a person or entity, across a set of agreed-upon attributes. This specific process raises questions surrounding what organization should be collecting the data, and

the means by which personal identity information should be kept up to date and relevant. In the case of financial services, organizations such as KYC.com have established clearing houses of identity data to enable customers to conduct more efficient KYC checks.

However, identity creation still presents a looming problem in many parts of the world. It is estimated that 1.1 billion people globally currently lack an officially recognized identity, and around 375 million adults in developing markets are unable to access financial services due to lack of required identity documentation.[4] In the absence of reliable government infrastructure to register people born or companies formed within a country's borders, often there is a void for other mechanisms of identity creation.

3.2 Verification

The second step of the identity lifecycle, verification, has been referenced in the previous section as especially problematic for financial institutions. Identity verification refers to the process of confirming at least one attribute of an individual or entity, either through self-attestation or third-party confirmation. Sometimes referred to as "identity proofing," verification looks to prove that trusted credentials or attributes are connected to the intended individual.

Identity verification is frequently discussed in the context of financial services: KYC and AML protocols rely on effective verification procedures. Financial institutions must rely on a combination of user-provided information and third-party attestations (the government may attest to a citizen's social security number, a utility company to a customer's address) to prove that prospective customers truly are who they say they are. Only with a verified identity can financial institutions initiate trusted transactions.

3.3 Authentication

Identity authentication, the third component of the identity lifecycle, is the process of determining that an organization is transacting with the same individual or entity iteratively over time.

The classic example of authentication in the digital age is the ubiquitous username and password. When a customer logs into their bank account, their financial institution needs to know that the person accessing the account is, in fact, the account's owner. Logging in with a username and password is one means of indicating to the institution that it is dealing with the same person in each transaction. Note that authentication does not necessarily require verification – that is, for standalone authentication procedures the particular identity attributes of the entity being authenticated are not being examined, as long as the authenticator can confirm that the entity is identical across transactions.

There are multiple additional methods for conducting digital authentication procedures, some of which can involve multiple factors to enhance security and reliability. For example, combining something the user knows (like a password), with something the user has (a device or credential), something the user is (a biometric marker like a fingerprint or iris scan), or something the user does (behavioral biometric analysis).

To that end, security and user experience are the twin primary concerns with most authentication procedures, and the two are often inversely related in legacy systems. As people access more disparate services online, it is increasingly convenient for them to reuse passwords across service providers. Various studies report that between 70-90% of consumers reuse passwords. This erodes security for individuals whose personal information is more likely to be compromised, and leads to enormous costs for institutions in the form of theft or compliance fines.

More secure technologies for digital identity authentication exist in various stages of development (multi-factor authentication, biometrics, and behavioral analytics, to name a few), but can be less convenient for users and difficult for entrenched institutions to adopt. Advanced biological and behavioral biometrics have also tended to provoke privacy concerns in some markets. Improving both security and user experience simultaneously is the primary driver for much of the technological innovation for this use case.

3.4 Authorization

Authorization is the process of determining what users can and cannot do based on their digital identity. It typically takes a combination of verification and authentication events to grant a user permission to perform certain actions. For example, after logging into their Netflix account, a customer will be granted access to streaming services based on their status as a paying member. However, if that user travels outside the U.S., they may not be authorized to view certain content based on a change in their location, a core identity

[4] ID4D, 2017, "Making everyone count," Identification for Development, World Bank, http://bit.ly/2FGgYxY

attribute in this transaction. From a service provider's prospective, effective authorization procedures involve robust internal process flows built on a foundation of accurate verification and authentication processes. A trend in authorization has been to move from role-based (a defined set of static permissions) to attribute-based (a more dynamic set of permissions).

Authorization fundamentally requires flexibility, as both roles and attributes change frequently and users authenticate (or fail to authenticate) into systems on a regular basis. Failure to accurately monitor key identity attributes could lead to illegitimate access of sensitive information or costly services. At the same time, however, it is an untenable burden for companies, in terms of both cost and security, to undergo continuous identity verification for all customers in order to ensure roles and attributes have remained constant for authorization purposes.

3.5 Federation

Identity federation is often the last step of a given digital identity lifecycle. Federation is the process of linking a digital identity or specific identity attributes across multiple distinct systems, or even across different service providers.

Establishing methods to execute federated identities has become increasingly attractive as the ratio of online to physical interactions increases. The most visible manifestation of identity federation are "single-sign on" (SSO) configurations by which a user can access multiple service providers through a single authentication process. Depending on the nature of the transaction, a service provider can federate an entity's verified, authenticated, or authorized identity – any of those functions can be shared. Identity federation is one approach toward reducing the burden of duplicative procedures outlined above.

In the world of access to social platforms, Facebook, in particular, has become a common federated identity service provider. Through the platform's 0Auth 2.0 capability, developers of digital services can connect their platforms to Facebook, with Facebook validating their login and then providing an agreed set of personal data to that application. In this particular case, maintaining that information is largely the user's responsibility. In other applications of federated identities, however, the consequences of stale, incorrect, or improperly shared data can have severe consequences.

Securing personally identifiable data is a challenge within one siloed service provider, and that problem only multiplies as identities are shared across institutions. With multiple interconnected accounts, the difficulty of achieving illegitimate access decreases while the incentive for doing so rises dramatically. Data ownership and consent also becomes an issue with federation — users are often not aware of how their identity data is used across accounts, and lose control of who can access their data and for what purposes.

Underpinning these five distinct identity building blocks are industry-, sector-, or jurisdiction-specific sets of identity standards. Standards concern an agreement between organizations and entities that are involved in a transaction with regards to what attributes of a customer are sufficient to create a trusted digital identity, and how that digital identity can then be verified, authenticated, and federated. An increasing number of government institutions and private sector consortium groups are advocating for open identity standards to bolster security, privacy, and user experience across identity use cases. However, identity standards can be very different depending on what is being transacted or what service is being accessed, and many are still evolving as technologies develop.

4. IDENTITY CHALLENGES FOR FINANCIAL INSTITUTIONS

Despite the unprecedented technological development and innovation in the financial services sector, financial institutions still face a number of considerable challenges in integrating digital identities into their services across these five identity lifecycle stages. Digital identity issues in the financial services space fall into a few major categories:

- **Administrative costs,** including manual verification, legacy record storage, and customer service costs.
- **Service delivery challenges,** including inability to tailor service offerings, inaccurate pricing, and customer exclusion.
- **Risk and compliance challenges,** including escalating KYC and AML costs as well as navigating new regulatory regimes like the E.U.'s General Data Protection Regulation (GDPR) and revised Payment Services Directive (PSD2).
- **Theft and fraud,** including escalating new account fraud, account takeover, and synthetic identity fraud.

Given the roadblocks currently in place, progress in this area has been slow, though there are opportunities to address each of these challenges through effective identity ecosystem development. Consider that under the current systems, customers must re-share the same identity information every time they want to do something as basic as opening a bank account or applying for a credit card. As improvements in digital identity become more universal, these additional steps should become a thing of the past, as banks gain access to decentralized and verifiable forms of identity that allow them to accept each other's approvals.

4.1 Administrative costs

Incomplete, ineffective, or outdated identity systems represent a significant cost to financial services providers and customers alike. When onboarding a customer, initial identity creation, verification, authentication, and authorization processes require individuals or entities to present physical documents or conduct in-person visits. As discussed above, manual verification of physical credentials represents a substantial investment of time and resources. The average cost of an in-person transaction is around U.S.$4.25, while mobile transactions reduce that figure to only U.S.$0.10.[5] Where fully digital identity authentications can take place using voice confirmation or biometric scanning technology, for example, transaction costs can be greatly reduced.

In the U.K., for example, 25% of financial services applications are abandoned by customers due to friction created by KYC.[6] Steps such as login or payment verification present challenges across a range of industries, but they are particularly problematic in banking. For example, roughly 30% of calls to bank call centers are requests for account access.[7] It's estimated that each of these calls can cost a company around U.S.$25 – a princely sum for basic customer service, all over something as simple as a forgotten password. In this way, a lack of digital identity represents a direct cost inefficiency to service providers and consumers alike.

4.2 Service delivery challenges

Financial services organizations can also gain advantages by analyzing customer identity data they have already collected and are not yet using. This is because data about customers has been traditionally housed in the individual, transactional systems outlined above, and are typically not well integrated across organizational divisions. This is known as a data silo, where an abundance of information about a customer is available, but is operationally unusable. Without the ability to intelligently interpret the data already collected, banks are unable to connect the dots and compose an integrated view of the customer.

[5] Fiserv, 2016, "Mobile banking adoption: where is the revenue for financial institutions?" https://fisv.co/2oEqLME

[6] Meola, A., 2016, "E-Commerce retailers are losing their customers because of this one critical mistake," BusinessInsider, March 16, http://read.bi/1puwynf

[7] Accenture, 2013, "The future of identity in banking," https://accntu.re/1S3FaHb

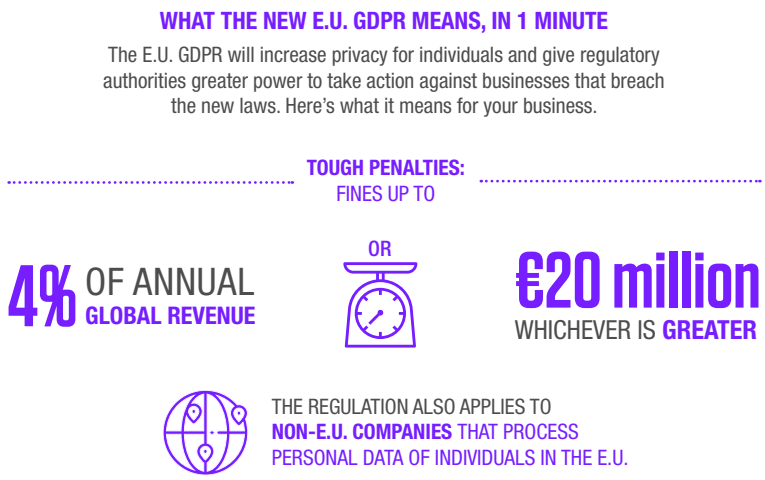
This poor management of customer identity can lead to a wide array of missed opportunities in service delivery. Consumer pricing, for example, is key in the financial sector. Here, banks could build targeted propositions to customers with pricing that reflects that customer's relationship with the bank. Institutions could also draw from rich transaction history data, which offers key insights into their buying habits. Banks, however, are not usually in a position to do this. For example, a bank would not want to price a personal loan independently of a mortgage – instead, they are likely to provide a competitive price that reflects the potential share of the consumer's available funds. The result is a fragmented situation, where each product line of a bank interacts with a customer as if it is the customer's first time doing business with that bank. From the customer's perspective, this is an impersonal and inefficient way of doing business. This issue was recently highlighted by Nomis, which found that banks can have over 300,000 pricing points across as many as 300 retail locations. And customers have taken notice – a Capgemini survey found that just 37% of customers believe banks understand their needs and preferences adequately.[8]

More broadly, lack of digitization throughout the consumer lifecycle, including reliance on physical identity creation and verification channels, excludes millions of potential financial services consumers. In the financial services sector alone, digitization could bring an additional 1.6 billion customers from developing markets into the formal economy, creating U.S.$4.2 trillion in new deposits and U.S.$2.1 trillion in new lines of credit.[9]

4.3 Risk and compliance challenges

The increased complexity of finance in the digital age has also led to an array of new issues related to compliance and risk, many of which have their roots in identity processes. Consider cross-border payments, where user verification can present a significant challenge. Correspondent banks in western financial hubs, such as New York or London, may be asked to handle payments from counterparties with accounts from countries where identity standards policies are less strict. It would be impractical for a bank to perform due diligence on each and every counterparty and transaction. As a result, institutions instead rely on algorithms intended to track payment flows and flag suspicious behavior. Unfortunately, in practice, these methods are not particularly effective, determining the probability of fraudulent activity without certainty. It is

Figure 2: E.U.'s GDPR



Source: IT Governance

estimated that financial institutions spent more than U.S.$8 billion on AML efforts in 2017, and it's expected that those investments will grow by 9% in 2018.[10]

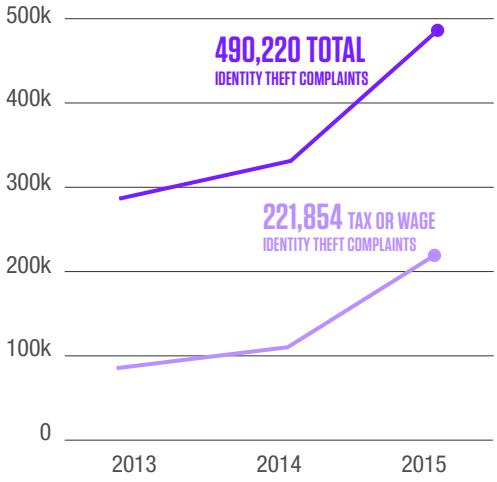
These identity challenges may become even more acute in 2018 and beyond. There are larger challenges on the horizon, including upcoming regulations that govern how data about customers can be gathered, used, and stored. For example, the General Data Protection Regulation (GDPR) (Figure 2) will place significant restrictions on the lifecycle of consumer data used by financial institutions, resulting in stiff penalties for noncompliance – up to 4% of global revenue or €20 million, whichever is greater. GDPR will also require that data collected about customers be commensurate with the product that the data is collected for. As a result, financial institutions will not be able to indiscriminately build up datasets on customers in anticipation that this information could potentially be used at a later point in time. GDPR also includes a "right to be forgotten" clause that will require financial institutions to delete all data concerning a customer when requested. Since many large banks have data trapped in silos, the lack of a centralized repository of customer information will only compound problems for these institutions. The global fortune 500 will spend an

[8] CapGemini, 2017, "Big data alchemy: how can banks maximize the value of their customer data?" http://bit.ly/2oKmZ4K

[9] Manyika, J., S. Lund, M. Singer, O. White, and C. Berry, 2016, "How digital finance could boost growth in emerging economies," McKinsey & Co., http://bit.ly/2z9Tpcm

[10] PwC, 2018, "Pulling fraud out of the shadows: the biggest competitor you didn't know you had," https://pwc.to/2sKL1xF

Figure 3: Identity theft



Source: Krebs on Security

estimated U.S.$8 billion to become GDPR compliant, and digitizing the identity management lifecycle will be a priority to stay in line with this new data protection regime.[11]

GDPR is not the only transformational regulatory regime reshaping the financial services sector, however. The Revised Payment Services Directive (PSD2) in the E.U. is aimed at modernizing European payment infrastructure and spurring innovation in payments and financial services. Its key provisions include a move toward "open banking," wherein existing financial services institutions must make consumer account information available to third parties (including new fintech players). PSD2 will lower barriers to entry for non-traditional financial players. This means that traditional financial institutions will no longer be able to rely on data access as an exclusive competitive advantage, and will be forced to innovate based on trusted consumer experience.

4.4 Theft and fraud

Identity management efforts with limited resources also inevitably lead to bad actors slipping through the cracks. Financial institutions are well aware that the vectors for theft and fraud evolve as quickly as the technological tools to contain them.[12] 15.4 million Americans were the victims of identity fraud in 2016, with losses totaling U.S.$16 billion.[13] Worldwide identity theft costs are estimated to be at least U.S.$221 billion.[14] Currently, an estimated 1 in 9 digital account creation attempts are fraudulent, as are around 1 in 20 digital login attempts.[15] Traditional identity processes are simply insufficient to contain digital threats.

The problem of synthetic identity fraud is a particularly urgent symptom of the existing identity problem. Fraud committed by consumers using synthetic identities – that is, exploiting weak identity creation and verification processes by combining a series of legitimate attributes to form a new, fictional identity – is growing. Up to 20 percent of defaulted credit card debt may already be the result of synthetic identity fraud, and the technique already costs businesses more than U.S.$6 billion annually.[16] For financial institutions, the problem is exacerbated by the lack of integrated customer records as discussed above.

5. IDENTITY AND TRUST IN FINANCIAL INSTITUTIONS

Beyond these direct revenue, compliance, and fraud considerations driving financial institutions to implement digital identity processes, identity is also a foundational component of trust and safety. The ability to execute trusted, secure transactions is a core mandate for legacy financial institutions looking to maintain market share as the landscape of alternative digital and mobile financial service options continues to expand. Connected customers, fatigued by 2017's unprecedented personal data breaches and able to select from a growing array of innovative financial products, make their choices based on trust. For traditional financial institutions, this means that trust is a core product offering – as quantifiable and impactful as any credit vehicle.

Broadly, trust and safety[17] refers to the full set of business values and practices that increase participation in and engagement with a digital ecosystem by reducing the risk of harm, fraud, or other criminal behavior toward an individual or organization and its reputation. Trust also requires that institutions have proper recourse mechanisms in place for redressing the damage of adverse events when they occur. By establishing a

[11] IAPP and EY, 2017, "2017 privacy governance report," http://bit.ly/2GVjgsl
[12] For more information on common identity-based vectors for theft and fraud see OWI, 2018, "Personal data management fundamentals," One World Identity, January 30, http://bit.ly/2s7i0Qq
[13] Pascual, A., K. Marchini, and S. Miller, 2017, "2017 identity fraud: securing the connected life," Javelin Strategy, http://bit.ly/2mYmaDi
[14] Carbajo, M., 2013, "How to prevent and detect business identity theft," U.S. Small Business Administration, January 9, http://bit.ly/1E16rsR
[15] ThreatMatrix, 2017, "Cybercrime report 2017: year in review," http://bit.ly/2oGlynN
[16] Auriemma Consulting Group, 2017, "Synthetic identity fraud cost banks $6 Billion in 2016," BusinessInsider, August 1, http://read.bi/2F90S27
[17] OWI, 2017, "Commitment issues: trust & safety through the digital fog," One World Identity, October 30, http://bit.ly/2GRx4E9

basic threshold of trust, a stakeholder will choose to participate in a particular digital ecosystem. Maintaining a sense of safety ensures nothing goes wrong when participating in that ecosystem.

Effective digital identity processes underpin the trust-building financial institutions must prioritize. They need to do so with two distinct constituencies: customers and regulators.

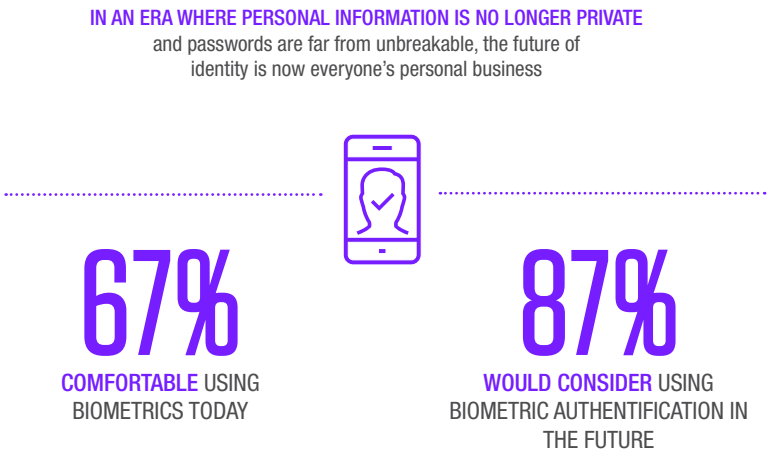
With customers, effective digital identity processes have the potential to minimize friction in user experience and enhance data security, both key pillars of trust and safety.[18] For example, new research indicates that customers more likely to trust financial institutions that use advanced technology like biometrics for identity verification and authentication.[19] In fact, over 40% of consumers would refuse to use a digital financial service that is not secured by some sort of biometric authentication.[20] Lack of user familiarity has often been cited as a primary obstacle in the adoption of new authentication technologies, but that is increasingly untrue. Customers now carry advanced fingerprint and facial recognition technology in their pockets, and are now increasingly demanding digital identity verification, authentication, and authorization as part of financial transactions. Simple passwords and traditional knowledge-based authentication mechanisms are, rightly, no longer as trusted.

Building trust with regulators is a related, but substantially more complex process. 2018 may prove to be a turning point for the regulation of personal data in markets around the world, and financial institutions must be proactive in building their identity data stewardship infrastructure to avoid crippling fines or sanctions under GDPR, PSD2, the Chinese Cybersecurity Law, or any of the other emerging data governance regimes under which they may fall. Each of these statutes requires financial institutions to have thorough knowledge of the personal identity data they collect, the business processes for which it is used, and the manner in which it is stored. Robust digital identity processes throughout the consumer lifecycle – from onboarding through the termination of the business relationship (which, under GDPR, may require the destruction of all personal data) – are a requirement for modern compliance.

When that trust is lost, financial institutions face potentially disastrous financial and reputational costs. On the consumer side, customers have exhibited decreasing trust in traditional banking institutions year over year.[21] Nearly 90% of customers say they will abandon a service provider that does not manage their personal identity data responsibly.[22] Overall, the average cost of reputation damage of lost trust due to identity data compromise ranges from U.S.$184 to U.S.$332 million.[23] In the case of Wells Fargo, for example, illegitimate use of personal data directly impacting around 3% of customers ended up costing the company an estimated U.S.$99 billion in deposits, and almost a third of existing customers reported looking elsewhere for banking services.[24] Quite simply, lack of trust costs banks customers, and digital identities are necessary for trust.

The good news for financial institutions, however, is that trust-building provides a ripe opportunity for innovation and differentiation. Currently only about a third of customers perceive significant differentiation between financial services providers based on product offerings alone.[25] For that reason, improving user experience and security through reliable and frictionless digital identity

Figure 4: The future of identity



Source: IBM Biometrics

[18] OWI, 2018, "Five pillars of trust and safety," One World Identity, January 5, http://bit.ly/2oFp3ut
[19] Sposito, S., 2018, "Two-factor authentication: even Google struggles to enroll users," Javelin Strategy, February 5, http://bit.ly/2FaPH98
[20] Security, 2017, "Consumers trust biometrics for mobile banking and payments," May 6, http://bit.ly/2HU98RO
[21] EY, 2017, "The relevance challenge: what retail banks must do to remain in the game," https://go.ey.com/2ihm5sI
[22] Kawamoto, D., 2017, "Consumers don't trust businesses can protect their data," DarkReading, http://ubm.io/2zcit6k
[23] Ponemon Institute, 2011, "Reputation impact of a data breach: U.S. study of executives & managers," http://bit.ly/2CQ8OzO
[24] White, G. B., 2017, "The toll of Wells Fargo's account scandal," The Atlantic, April 19, http://theatln.tc/2Fa9eXf

creation, verification, and authentication procedures can itself be a differentiator in the increasingly crowded market for digital financial services. Effective digital identity processes, and the trust they engender with customers, are a competitive advantage that financial institutions should explore.

6. FINANCIAL INSTITUTIONS AND THE FUTURE OF DIGITAL IDENTITY

Financial institutions are fundamentally identity-centric institutions. For trusted transactions to take place in the digital economy, institutions must invest in constructing effective digital identity infrastructure throughout the customer identity lifecycle. While this will require significant attention to mitigating the identity challenges outlined above, it also means that financial institutions are uniquely positioned to support the development of digital identity ecosystems across sectors.

Traditionally, the financial institutions have been a key component of an identity architecture from the perspective of enabling merchants and customers to confirm that they are who they say they are. For example, in credit card networks, both merchants and customers are validated by banks.

However, financial services landscape is increasingly moving toward a less tightly-bound ecosystem. For instance, the frequency of cross-border transactions is increasing, involving customers and client organizations who are members of non-domestic banks with different verification standards. Peer-to-peer lending organizations and non-depository payment providers are proliferating, such that there may be no traditional banks involved in a financial transaction. Gaps in the existing digital identity structure are becoming a significant constraint, particularly as fintech organizations continue to enter and disrupt the market.

This is where financial institutions have a potential role to play. These institutions are trusted with processing large amounts of persona data, and have been performing an identity broker role in some form for some significant time. Financial institutions, therefore, have the opportunity to offer identity verification, authentication, and federation services to organizations both within the financial services sector, and even in cross-sector use cases.

Recent research has already highlighted the extensive potential for financial institutions to facilitate identity services in both public and private sector interactions.[26] Indeed, in some markets, new nationwide identity infrastructure layers are being constructed driven primarily by financial institution participation.

The U.K.'s GOV.UK Verify system, for example, allows users to access public sector services online after their identity is verified by a private company of the user's choice, like Barclay's or Experian. In Canada, SecureKey Concierge follows a similar model with several financial institutions serving as identity providers for citizens to access dozens government services. Sweden's BankID platform facilitates identity services for 2 billion transactions per year.[27] BankID has recently integrated next generation identity verification and authentication mechanisms based on behavioral biometrics to minimize reliance on passwords. Six of the country's largest banks also cooperatively launched a common mobile payment app, Swish, in 2012, building on BankID's functionality.

Exporting identity services has already proven to be a successful endeavor for traditional financial institutions in these markets. Institutional liability and trust risks remain, however, as this business model continues to mature. If Bank A relies on Bank B's attestation of a customer's identity, for example, and that initial attestation is later determined to have been insufficiently thorough, Bank A could feasibly have recourse to pursue damages for any fraud committed in some jurisdictions. At a time when financial institutions are receiving unprecedented fines for lax customer due diligence, this could be an area in which some organizations have a low appetite for risk.

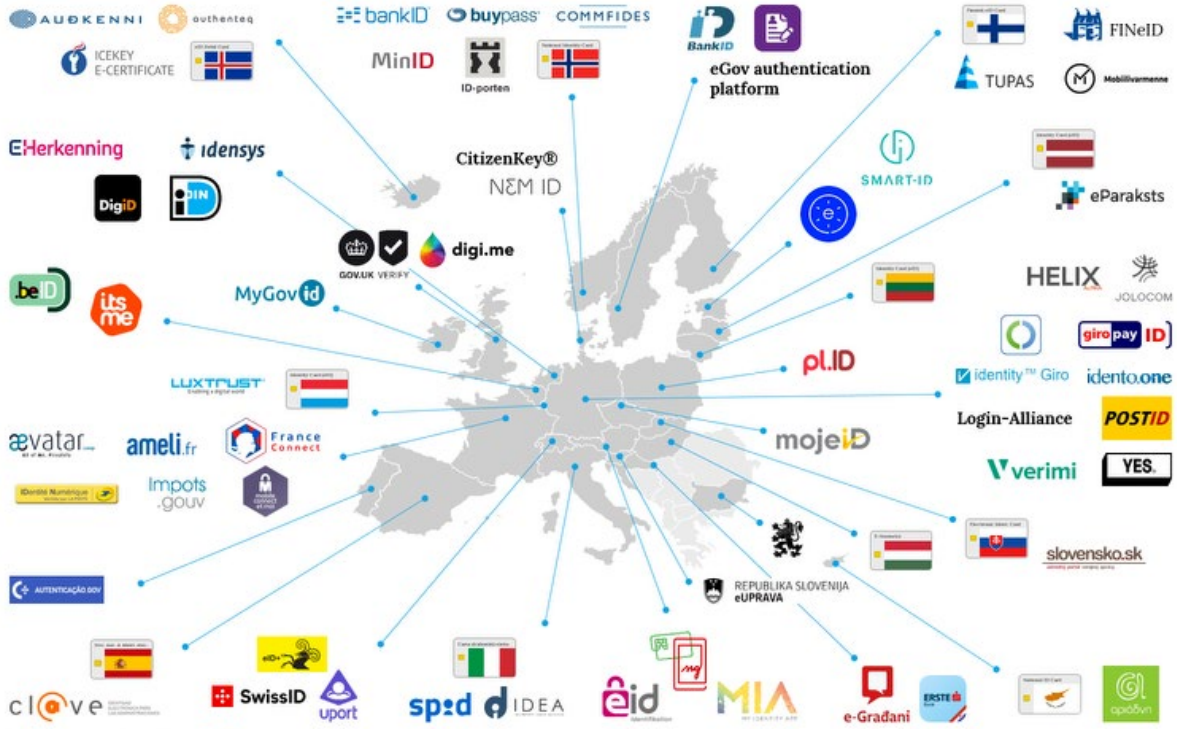
Nevertheless, as legacy banks struggle to maintain relevance and market share in an increasingly decentralized financial services sector, digital identity, and the consumer trust it engenders, could itself be a profitable service offering in the connected economy.

[25] EY, 2017, "The relevance challenge: what retail banks must do to remain in the game," https://go.ey.com/2ihm5sl

[26] World Economic Forum, 2016, "A blueprint for digital identity: the role of financial institutions in building digital identity," http://bit.ly/2aOblg1

[27] Metzger, M., 2016, "ISSE 2016: The four models of digital identity," SC Media, November 23, http://bit.ly/2HVKAr0

Figure 5: Key identity initiatives within Europe



Source: asquared

7. CONCLUSIONS AND A LOOK AHEAD

Identity – of customers, client organizations, and partner entities – is at the heart of the financial services industry. Without effective identity processes, clients and regulators lose trust, financial institutions lose money, and legacy institutions lose out to the alternative financial services players emerging as part of the fintech wave. But, within the identity challenge lies an immense opportunity for financial institutions to build the infrastructure for future cross-sector digital identity ecosystems. A few core lessons will help financial institutions adapt to the reality of the connected economy and lead the evolution of digital identity:

- **Legacy, paper-based identity processes are expensive and unreliable.** Traditional identity creation, verification, and authentication procedures in particular are costing financial institutions not just money, but also time, trust, and competitive edge. Innovative identity solutions, including advanced authentication mechanisms like biometrics and behavioral analytics, improved internal data stewardship, and enhanced digital and mobile service offerings, can significantly reduce administrative costs, bolster security, and improve customer engagement.
- **Effective digital identity systems are necessary for institutional survival.** In today's digital economy, trust in traditional financial institutions is falling, and customers are less likely to perceive differentiation between banks based on product offerings alone. A more educated generation of financial consumers will choose to interact with financial institutions they trust. Robust digital identity processes build trust and safety with users and regulators by enhancing user experience and security. Both will be required for banks to stay relevant.
- **With new regulatory regimes, data access is no longer a competitive advantage, but trusted identity services can be.** 2018 will be a year of fundamental shifts in the regulatory landscape. Barriers to entry for innovative fintechs are falling, but the standards for collecting, sharing, and storing identity data are more stringent than ever. Banks are no longer the sole custodians of customers' economic destiny. Establishing trust through frictionless and secure digital identity processes will be key for customer retention.
- **Financial institutions are uniquely positioned to underpin digital identity ecosystem.** As developing identity ecosystems like those in the U.K., Canada, and the Nordic countries have demonstrated, financial institutions are uniquely positioned to drive the development of digital identity ecosystems that extend across the public and private sectors. Demand for effective digital identities is growing in nearly every consumer-facing industry, and financial institutions can play a key role in providing the identity services as the foundation of trusted transactions for years to come.

Setting a standard path forward for KYC

ROBERT CHRISTIE | Principal Consultant, Capco

ABSTRACT

Customers have a good reason to be upset with banks over their KYC processes, which tend to be complicated and costly. Given the pressure and timelines from regulators, it is understandable that banks have struggled to make KYC customer-friendly. With new technologies becoming rapidly available, now is the perfect time to set a new standard for eKYC solutions that would make compliance fast and cost-effective to implement. However, there is a key dependency that needs to be considered before a global solution can be delivered. This article provides some recommendations on how this could be achieved.

1. INTRODUCTION

Since 2001, regulatory bodies across the world have introduced a wide array of regulations targeted at the opening and maintenance of bank accounts by individuals and corporates. This increase of regulatory scrutiny had arisen from increased concerns over money laundering and the use of the global banking system to finance terrorist activities.

Under these new regulations banks are more accountable for detecting and preventing money laundering. This has pushed them to develop new processes and systems, hire extra compliance staff, closely monitor transaction activity through accounts, and report any suspicious activity detected. However, as banks struggle to keep up with new regulations and implement procedural and technical changes to support them, there has been an unintended impact on the banking customer who struggles to understand information requests and comply with account opening and maintenance requirements.

As these impacts on the customer mount, and the costs to become compliant increase for banks, greater pressure is being placed on banks to develop solutions that will facilitate detection of money laundering. Developing a quick solution is however, a significant challenge for the banking industry.

2. IT ALL STARTS WITH DUE DILIGENCE

Know Your Customer (KYC) is a process whereby a financial institution verifies the identity of an account holder and understands the purpose of the account, otherwise known as performing Customer Due Diligence (CDD).

CDD first became formalized under the "40 recommendations" issued in 1990 by the Financial Action Task Force (FATF), where guiding principles on how to conduct CDD were defined for banking regulators around the world. In 2001, in response to weaknesses in how banks were implementing KYC processes to support customer due diligence, the Basel Committee published "Customer due diligence for banks," which aimed to strengthen this critical component of anti-money laundering and counter-terrorist financing.

Regulators responded to both the FATF recommendations and the Basel Committee with guidelines and regulations of their own. For example, the U.S. Patriot Act, which introduced the Customer Identification Program (CIP), aims to establish compliance standards for U.S. banks to follow when identifying the identity of an account holder. In simple terms, each bank must have a sufficient degree of certainty about the identity of the account holder and perform the necessary due diligence to verify that the information is true and correct. CIP programs now form the core of most AML regulations and KYC policies around the world, with each regulatory body enacting its own form of the guidelines.

The general requirements of CIP specify that financial institutions must collect documentation that prove the account holder's identity (such as a government issued identity card) in order to validate the exact name, nationality, and date of birth of the individual. This information is then used to ensure that the account holder has been clearly identified, and in the event of concerns raised over the use of the account, the bank will know exactly who to hold accountable.

Once documentation is provided by the customer to support the CIP requirements under the KYC process, and it has been evaluated for clarity, certainty, and risk, the bank should have a clear understanding of the customer's identity. Should any risk items have been flagged during the account opening, the KYC process would prevent the account opening from proceeding until safeguards had been put in place to mitigate the risk, or possibly even prevent the account from being opened in the first place.

On simple review, the information that is captured by a KYC program to satisfy CIP requirements should be easy for any customer to provide, and straightforward for any bank to collect and store. For example, government issued identity documents help to verify the identity of individuals, company registration certificates verify the formation of a company, board resolutions provide the necessary mandates for account opening, and organizational structures identify who has control and influence over account activities. These are all standard documents that any individual or company should have readily available to provide on demand.

Unfortunately, despite the simplicity of the request, there are underlying challenges that both the customer and the bank must overcome before the KYC process can be completed and the account be opened. What may appear to be a simple request on the surface can actually unearth many complexities that both the customer and the financial institution must resolve together.

3. THE NUANCES OF REGULATIONS COMPLICATE THE EFFORT

When taking a closer look at CIP requirements, there are significant nuances that challenge banks as they attempt to build KYC processes and procedures. In an effort to build a KYC framework that can accommodate each and every customer, the exceptions to the norm often derail the efforts to develop a simple process. As the bank attempts to build a single KYC process that accommodates a variety of customers, the process becomes increasingly convoluted and difficult to implement.

Take, for example, a simple requirement to present a document as proof of identity. This basic requirement immediately raises a myriad of questions and concerns for the bank. Is the bank required to be an expert in every government issued document worldwide? Is the bank responsible for ensuring that the document presented is valid? How does the bank know that the identity document is issued by a trustworthy body or official government agency? And if the customer is not physically present, how does the bank know that the document being presented is truly the individual opening the account?

These challenging questions get further complicated when the account beneficial owner does not have the documentation specified under the KYC requirement. For example, a U.S. citizen is not required by law to possess a government issued identity document, which is a standard requirement under KYC for many countries. If a U.S. citizen wants to prove identity and nationality, a birth certificate may be the only option available. However, if that individual has an account outside the U.S., a birth certificate may not be sufficient to satisfy local regulations, since it does not have a photo image of the individual. Again, the bank is placed in a difficult position of not knowing whether a legitimate document can be accepted as proof of identity, and the customer may truly have no other options to consider.

Lastly, customers may be very uncomfortable providing some forms of identity documents due to concerns over privacy. Government issued identity documents are generally accepted as means of proving identity, but in many countries they are seen as confidential documents. As the CIP requires that a certified true copy of the document be provided to the bank, the customer has the additional worry that the identity document copy is safeguarded against theft or intrusion.

4. THE ONLY CERTAINTY IS THAT ACCOUNT OPENING TAKES TIME

Unfortunately, the KYC process can quickly start to unravel as more nuances are discovered and the compliance more challenging. Decision-making to resolve the nuances takes time, as compliance officers are brought into the discussion and review to negotiate with business stakeholders. As the dialog evolves, especially around complicated situations where more risk is at stake, the customer must wait for a resolution.

"As the bank attempts to build a single KYC process that accommodates a variety of customers, the process becomes increasingly convoluted and difficult to implement."

This obviously impacts the customer, who needs to wait until the situation is resolved before the account can be opened. Not long ago, an account (even a business account) could be opened within two working days. Today, banks are reluctant to quote timelines to prospective customers because they know that the process could drag on for weeks and sometimes months.

Corporate clients are particularly susceptible to these delays in account openings, where there is often a genuine time sensitivity to a transaction that can impact the success of their business. These delays can have significant impact on a business, especially a new business that may be growing quickly and needs to process transactions in a timely manner to build trust with business partners.

And, it is not only the customer that is losing business, the bank itself is also losing revenue opportunities. The longer the client is left waiting for the account to be opened, the more expensive the account opening becomes and the greater the loss of revenue opportunity. Despite these losses, both sides are equally helpless and must endure the challenges together in the hope that the impact is not too great.

5. CUSTOMERS HAVE NO CHOICE BUT TO COMPLAIN

Many bank customers complain about the tedious process and sometimes invasive lines of questioning that accompany the KYC process. As the customer is driven through the KYC process, the mounting requirements seem impossible to fulfill and become obstacles in opening (or maintenance) of the account. Weeks, and even months, can pass by as issues encountered during the KYC process are escalated for resolution by a compliance manager who may be overwhelmed by the volume of questions or simply needs time to consider the situation.

If banking was an industry competing with other industries for the same market segment, they would fail simply due to customer dissatisfaction. From the outside, it appears that the banking industry holds its market hostage and is dragging its feet on how to become more customer service oriented with its KYC process. The only recourse a customer has is to complain and hope that their voice is heard over the others that are also voicing their frustration.

But, where does the fault lie, with the bank or the regulator? Both are probably to blame. The regulators have mandated KYC requirements that do not consider the variety of challenges faced in implementing them. Meanwhile, the banks have struggled to devise programs, build systems, and educate customer service teams in how to handle the variety of situations that can occur during the KYC process.

Needless to say, both banks and regulators have recognized these faults and are making efforts to improve the client experience.

6. THE DILEMMA FOR BANKS

To be fair, banks are aware of the negative impact that these requirements have on their customers and are very concerned about it. However, they are caught in a dilemma: should they take the time to develop a client friendly KYC process that will take considerable effort and resources to implement and manage, or risk customer satisfaction with a KYC process that is quick from a regulatory approval perspective but does not provide a satisfactory customer experience?

Many banks have been caught in the crosshairs of the regulators by not having a compliance department that is well versed in global regulations. Finding compliance experts who can review regulations quickly and effectively and translate them into meaningful policies and procedures is a daunting task. Particularly challenging is the fact that regulators have given short timelines with strict penalties if the regulations are not met.

The fear for any financial institution is that they will fail an inspection by a regulator and lose their license to operate. Loss of operating license, or any restrictions on the business, is a blow that can ruin any bank overnight and cause tremendous harm to account holders and their respective businesses. Regulators clearly want to avoid this outcome as much as banks, so there is often a period of time given to the bank to become compliant. However, in the scramble to make the necessary changes there is always impact on customers who struggle to fully understand and accept the changes that are not always well explained.

Further complicating the dilemma is that the bank needs to implement compliance standards that are “global” and cover each jurisdiction in which the bank operates. Designing a framework that is global, simple to implement and enforce, and do so in a way that makes sense and with minimal impact to customers has been the largest challenge for all banks. Most have often erred on the side of caution by implementing overly rigorous “global standards” programs that are challenging to develop with procedures that are difficult and confusing for internal staff and customers to follow. Unfortunately, a major consequence of not implementing the correct KYC compliance program, or one that is too weak, again is to receive another fine or potentially lose a banking operations license, which is far too great a risk to consider.

For certain, the intention of applying regulations on financial institutions is not to cause harm or difficultly for account holders. Although it is difficult for most account holders to see the mechanics of these programs, the complexity of a compliance program that is equally uniform yet bespoke to certain types of customers is not a realistic approach towards solving the problem. When a bank has tens of thousands, or even hundreds of thousands, of accounts that may be impacted, there is simply no way to evaluate each account holder on a case-by-case basis within a realistic time frame or resource pool.

7. HOW TO EASE THE COMPLIANCE STANDOFF

Most banking customers have a story or complaint to share about their experience with KYC. The level of frustration is significant, with both customers and bank staff who are perplexed and annoyed with the challenges of being compliant.

Banks are not good at change, but they are making an effort. And customers are not good at compliance, but they are slowly accepting it and making it part of their business planning. This does not mean that both sides need to be content with the current situation. Opening a bank account should not take weeks or months, and customers should have the right to use their accounts legitimately without undue scrutiny while issues encountered during the KYC process are resolved.

The easy – and often stated – solution to simplifying KYC is technology. This is a fair statement, but it overlooks genuine questions and problems. Yes, technology will enable a solution and be a key component towards its success. However, the true problem is the lack of common data standards and protocols, which if agreed – and not only across the banking industry but also between regulatory bodies – could trigger a banking compliance revolution.

Take again, for example, the issue with identity documents and the challenges that banks face in evaluating and accepting them. The purpose of the identity document is to validate the name of the account holder, their nationality, and date of birth. For the most part, the identity information that each bank around the world is collecting under a KYC program is much the same. From a customer's prospective, this information is static and, therefore, needs to be validated only once, so that it can be accepted whenever needed by any bank worldwide.

Under a global verification model, the customer completes the identification verification process only once with a trusted third party (which could be a bank or an independent company). Verified details are then certified by the trusted third party with a digital certificate that is then linked to the encrypted personal data file. Upon request, the customer authorizes the bank to access the encrypted file which is then validated through a key exchange that confirms the right to access and the authenticity of the data. Upon confirmation from the trusted third party, only the necessary personal identity details are transmitted from the data file to the bank, which then feeds them into the back office KYC system.

There are many advantages to this model, which in various forms is becoming known as "eKYC." The customer only needs to complete the identity verification process once and retains ownership and control over their personal details. The bank no longer needs to review and validate identity details, saving it tremendous costs and resources, as well as reducing risk of error. Most importantly, the process can be achieved in seconds, as opposed to the days or even weeks that it currently takes to obtain certified true copies of documents and have them accepted by bank staff.

This secure technology is already in use today and is widely available. The problem is how to agree on a common standard data format and which third party will be the trusted authority to verify and certify the identity details. Before any bank could accept such an eKYC model, it would need to be sure that the data format is consistent and that the regulators have accepted the third party as the independent certifier of the data. But, as we look at how common standard can be developed for an eKYC solution, we must look back at the fundamental CIP requirements and how KYC identifies individual account holders.

8. INTRODUCING EKYC AS A STARTING POINT

Many regulators are approving the development of eKYC solutions, although what exactly this entails can differ between countries. Certain locations in Asia, such as Singapore, Malaysia, and India already have regulatory approval for eKYC, but each is taking a slightly different approach with development. Although regulatory approval has been provided, how exactly eKYC is to be accomplished has not been specified, nor have the expectations surrounding the underlying technology.

EKYC simply means performing KYC electronically, or without paper (as is the current practice). For example, instead of asking the customer to present certified true copies of identity documents on paper (such as a passport), the bank can accept a digital identification card that can verify the individual through biometric scanning (such as a fingerprint). Personal details are linked to the identity card, either in a memory chip on the card or accessible through a secure online channel, which are transmitted to the bank to support

the customer's KYC profile. The customer then needs to only present the identity card at the time of account opening in order for the bank to receive the details it needs.

The time (and cost) difference of using eKYC solutions to identity individuals is substantial. Providing a certified true copy of an identity document can cost up to U.S.$100 per copy. When you consider that the document is sent by post, time becomes a considerable cost as well. However, an eKYC solution that leverages an electronic identity card accomplishes the identity verification instantly and has virtually no cost once the hardware and software have been installed. For customers, this is a major improvement over the current situation.

9. SAME CONCEPT, DIFFERENT COUNTRY

Leveraging electronic identity cards is, therefore, the logical starting point towards building a full eKYC solution. Conceptually, the electronic identity card is providing the same information about the individual as a standard government-issued document, such as a passport: name, date of birth, nationality, and possibly birth place and current residential address. However, even with those basic details in mind, every country is taking its own route with electronic identity cards.

In Malaysia, the MyKad identity card is carried by all Malaysia citizens. This identity card contains a chip that stores basic personal information such as name, date of birth, place or birth, residential address, and most importantly, a digital copy of a fingerprint along with a photo image of the individual embossed on the card. By combining the personal information along with the biometric validation, the MyKad can provide all required information under a customer identification program (CIP) to satisfy KYC requirements, which the customer simply needs to present at account opening.

In India, a program managed by the Unique Identification Authority of India (UIDAI) has been developed to issue a unique 12-digit identity number, called Aadhaar, to all individuals. Upon opening of a bank account, the customer provides their Aadhaar number and then authorizes the UIDAI to release personal details through either a single-use password or biometric verification. The bank account is then linked to the Aadhaar, which further allows the bank to receive the personal details and be immediately updated whenever there is a change.

In both examples, a unique identifier number has been assigned to the individual. It becomes a single point to which personal details are attached through an electronic storage mechanism. The difference between the two identity cards is around the technology used and the means of transmitting and verifying the data. Whereas the MyKad stores details on a memory chip embedded in the plastic card that can be verified by a fingerprint scan, the Aadhaar transmits details from a database held by the UIDAI and then verified through a password. Fundamentally, the data is the same but

the underlying technology is different enough to make them unique eKYC solutions. Yet, both are part of their respective countries' strategies in adopting an eKYC solution for their local banking industries, which are already proving to be a significant success in reducing time and costs associated with account opening and maintenance.

10. ADVANTAGE – LOCAL BANK

However, it is those technical differences in the approach towards electronic identity cards that make developing a universal eKYC solution so difficult. Despite the progress at the local level, solutions that are universal and span across borders are still out of reach. Consequently, the advantage is currently with the local banks that operate exclusively (or majority) in their home country. Because the local bank's resources are focused on the local market, they are at liberty to invest in an eKYC solution that meets their local regulator's needs. For example, a local bank in India can comfortably invest in the hardware to support the Aadhar knowing that it is a government-approved standard for India.

For the global bank, however, this is a problem. Global banks have systems and infrastructure that are shared across locations and are difficult to customize to local country requirements without incurring significant costs. Building applications and technology that are bespoke to one country is only undertaken when it is absolutely critical to the operations of the business in that location or mandated by local regulators. Otherwise, the underlying technology must remain consistent in order to minimize costs.

Although not impossible, building eKYC solutions that meet each country's unique approach towards identity verification will be costly and difficult to maintain if governments continue to adopt their own approaches. At best, global eKYC solutions are years away from deployment as countries continue to explore and standardize the underlying technology of their identity card system. In the meantime, many customers are discovering that holding an account with a large global bank does not mean better service when it comes to KYC. In fact, fulfilling KYC requirements with a large global bank, even on accounts held locally, is time consuming and costly and unlikely to improve any time soon.

11. TAKING REQUIREMENTS TO THE CORPORATE LEVEL

It is important to recall that KYC and CIP apply not only to individuals, but to corporate customers as well. As companies are also considered legal persons that can be account holders, identifying the company as both its controlling party and beneficial owner is also a requirement under any KYC program.

Banks are more challenged to perform KYC on companies due to the complexity of the corporate structure and the number of parties that need to be involved in the KYC process. However, some countries have simplified the KYC process for banks by making it a part of the company registration. The German Commercial Register (Handelsregister) offers not only the legal name and address of the company but also the current details about the controlling officers and their respective identification information (as required under law), which are required under KYC.

Banks in Germany only need to obtain the company profile details from the Handelsregister to have most of the details that are required for the KYC profile. Considering that both the government and the financial institutions have a need to know, it makes sense that both can leverage the same "golden source" of information. The only downside is that the electronic verification of company profiles through the Handelsregister is only accepted in Germany, and should that company have accounts outside of Germany it will need to follow a traditional paper-based process to provide the same details.

12. USING BLOCKCHAIN TO UNBLOCK THE PATHWAY

In Thailand, the Ministry of Digital Affairs has recently signed a memorandum of understanding (MoU) with a digital firm to explore the use of Etherium blockchain technology to provide its citizens with a national digital ID. How exactly the blockchain technology will be applied to a national ID system in Thailand has yet to be announced, but it is a clear indicator that Thailand also intends to implement a secure system that will provide a unique identity number to each citizen. Again, similar to other countries in the Asian region, this technical approach lays the foundation for the development of an eKYC solution for Thailand.

There has been much discussion around the use of blockchain technology to facilitate KYC, and in many ways, it should be a part of the solution. To be clear, however, blockchain technology is used to build a historical record by documenting sequential events that are interlinked within the digital record. Each block of the digital record chain is a single event that is based (and dependent) on the block that preceded it. By examining the blocks in their sequential order, the historical record of the underlying subject (or object) can be clearly traced and audited.

Cryptocurrencies, such as Bitcoin and Etherium, have used blockchain to track the value of their currency by recording every transaction event within the lifecycle of the currency. Similarly, the entire financial history of an individual can be written into a blockchain that records each transaction as a historical event. From a banking perspective, this can be useful in helping to understand and analyze the customer and their financial position while ensuring that a truthful record can be consulted as needed. From a governmental perspective, personal details beyond name and date of birth can be recorded in a secure file that also records those changes. Use of blockchain will also help to ensure accuracy of the individual customer's data as they apply for banking services by providing a historical financial record that is reliable and can be leveraged immediately.

The other facet of blockchain technology is the "distributed ledger," which enables collaborative recording of the events into the blockchain. Distributed ledger means that the recording of blocks in the chain is shared between participants, thereby making the full blockchain history both recorded and accessible to everyone. Due to the distributed ledger approach, the blockchain record becomes a more comprehensive picture because it encompasses the recording of events from a variety of sources instead of just one.

Although there is significant value in having an accurate and comprehensive financial profile of the individual, we need to revert back to the immediate objective of KYC, which is to identify and verify the identity of the account holder. Blockchain provides a historical record, but does not verify the identity of the individual on its own. However, blockchain technology does serve the broader objectives of KYC, which is to understand the intended use of the bank account and whether it matches the historical profile of the individual or company. Consequently, it should be considered as part of an eKYC solution but not a solution on its own. Without identity verification, blockchain solutions for eKYC will not be effective. Meanwhile, leveraging blockchain technology to develop an eKYC solution in tandem with electronic identity cards is a logical step in reaching a target state KYC solution.

13. HARMONY MAY NOT BE PART OF THE MUSIC, YET

Unfortunately, time is costing the banks dearly with regards to supporting KYC requirements. The pressure to find quick-win eKYC solutions is immense, even if the target state solution has yet to be defined. However, eKYC is waiting on how each country will implement a national identity card system that is electronic and integrated with the local banking infrastructure.

From a banking industry perspective, there is a clear advantage in validating identity from a single golden source, such as a government body. Global banks are now faced with immense pressure to accommodate a variety of eKYC solutions to support different approaches adopted by governments. Unfortunately, the lack of harmonization in data formats, data sources, technical approach, and capture techniques is challenging global banks to develop underlying technologies to support them all.

Driving the issue further between global standards and local customization are the concerns of the customers themselves, who as citizens have rights to privacy protected by their governments. Each country has a different perspective on privacy, and what constitutes personal information protected under its laws. Estonia, for example, has taken a broad approach to capture a wide variety of personal information under a single e-residency program. Under this program, any person in the world has the opportunity to become an e-resident with a unique identification number that can used worldwide and applied to all types of personal details such as medical records and financial statements. In some countries such as the U.S., this would cause great concern over access to private information whereas in Estonia it is seen as helping people share personal details on a need-to-know basis.

Finding the right path through the privacy landscape is the fundamental challenge of a truly global eKYC solution. Each country will find its own direction that will satisfy its citizens. Unfortunately, that means a disharmonious approach that will continue to challenge banks to find common solutions. Blockchain may provide some relief here with its ability to provide masked data, but again it is not the first step and is still dependent on some form of nationality identification number.

14. CONCLUSION

The cost of KYC compliance has been exorbitant for banks, mainly due to the lack of technology to support the process and the need to follow paper-based processes to complete the work. Hiring compliance officers and analysts, building of new systems, and training staff have an annual price tag that is staggering, with costs reaching over hundreds of millions of dollars for the larger global banks. And this does not even consider the fines and penalties that banks must pay for being non-compliant. The cost for customers is also significant, but probably best measured in lost opportunity and frustration which could be argued as the greatest cost thus far.

Ours is a time of transition for the banking industry, so it should not be a surprise to anyone that these challenges exist. The important point is that all parties are doing what they can to simplify and comply with the laws of their host and other countries. KYC processes exist to protect everyone and stabilize the global banking infrastructure. A financial system where money laundering is rampant only leads to a society where everyone loses, so we can all agree that any regulation and effort to fight money laundering is paramount in the banking industry.

The unfortunate part of the story has been the slow adoption of tools to facilitate the KYC process. Regtech, as it has come to be known, is still in its early days, with technology companies small and large racing to bring tools to market but with no proven global solutions (as of yet), although there is proven success at the local level that can leverage electronic identity cards. Even though banks want to implement such tools on a broader scale, the lack of global standards is holding them back. However, once eKYC standards can be agreed by intergovernmental groups and country regulators, and a more uniform approach is adopted on how electronic identity cards are issued, the regtech market will be quick to deliver solutions and banks will be better equipped to implement them.

Meanwhile, banks are caught in the middle and waiting for standards to be developed and agreed not only between countries, but also within each country's legal system. Conceptually, we can see that eKYC will be a marriage between a national identity card system and blockchain technology. However, exactly which party is the holder of the privacy key in this equation, be it a government or a trusted third party, is fundamentally where the debate lies. Until that is resolved and standards are agreed, banks must wait before committing fully to any eKYC solution.

With an agreed set of standards, tools to support eKYC will find their way quickly into the marketplace. Ensuring that these tools comply with not only banking but also privacy laws will be critical in their success and adoption by customers. The technology already exists to build these tools, and many countries are already adopting them for their own citizens. Overcoming the standards obstacle will greatly simplify the KYC process. The central focus of banking can then shift away from the regulation that aims to protect the customer interest, back to customers themselves.

References

Basel Committee on Banking Supervision, 2001, “Customer due diligence for banks,” Bank for International Settlements, Basel

New Straits Times, 2017, “CIMB receives sandbox approval for paperless customer identification.”, November 23,

Financial Action Task Force, 2010, “FATF 40 recommendations,” FATF/OECD, Paris

Financial Action Task Force, 2017, “International standards on combating money laundering and the financing of terrorism and proliferation,” FATF/OECD, Paris

Gupta, R., 2018, “Why biometric identification is required for KYC in digital India.” BW Businessworld. February 12

Heller, N., “Estonia, The Digital Republic.” The New Yorker. December 18 & 25, 2017

U.S. Department of Justice, 2001, “Uniting and strengthening America by providing appropriate tools required to intercept and obstruct terrorism (USA Patriot Act) Act of 2001,” Government Publishing Office, Washington D.C.

White, B., 2018, “Thailand mulls Ethereum blockchain for KYC program,” Bitcoin Journal, February 19

E-residency: The next evolution of digital identity

CLARE SULLIVAN | Visiting Professor, Law Center and Fellow, Center for National Security and the Law Georgetown University, Washington D.C.

ABSTRACT

This article examines the next evolution in government-backed digital identity programs that enable public and private sector transactions by individuals. Estonia is the first nation to offer e-residency to individuals who are not Estonian citizens and who are not legally resident, or even physically present, in Estonia. The Estonian program is the first government-authenticated and operated, international digital identity program that enables remote-access international commercial transactions that range from establishing and operating a company, trading in goods and services, opening and operating a bank account, to buying and selling securities. While Estonian e-residency is designed to expand the economic base of Estonia beyond its geographical boundaries, and in that regard, it is successful and inspiring, its impact is much more profound and far reaching. In establishing e-residency, whereby anyone, based anywhere in the world, can do business and banking in Estonia, and then potentially in the European Union (E.U.) and elsewhere, Estonia is changing traditional approaches to immigration, residency, and international business. In effectively opening a new virtual domain, Estonia is redefining what it means to be a nation and a citizen in the digital era, and is challenging the very nature and scope of international commerce and finance, and of regulation based on physical boundaries.

1. INTRODUCTION

In December 2014, Estonia became the first nation to open its digital borders to persons throughout the world to become an Estonian e-resident. Estonia is the most advanced e-society in the world and is an acknowledged leader in technology innovation. The Estonian e-residency program is another example of the country's extraordinary vision and ingenuity. In launching e-residency, Estonia hoped it would be transformative and disruptive, and it has proved to be so.

The primary objective of the e-residency program is expansion of Estonia's economic base, which is limited by its geography and relatively low population of around one million inhabitants. Under the program, anyone, based, anywhere in the world, can become a virtual economic resident of Estonia. Estonia is a member of the E.U., so e-residency also facilitates broader commercial access to Europe.

When the program first launched, applicants had to go to Estonia to apply in person to become an e-resident and to open an Estonian bank account. Now, without ever setting foot in the nation, a person can apply to become an e-resident and obtain an e-ID issued by the Estonian government. Digital trust services, including electronic signatures and seals, and blockchain technology underpin the program to enable an e-resident to remotely access and use a range of Estonian e-government and private sector services. An e-resident is able to remotely perform a full range of commercial activities, including business and company registration and operation, banking (including funds transfers), buying and selling of real estate and other property, and trade in goods and services.

E-residents are subject to Estonian tax,[1] and e-residency does not operate as a tax shelter in relation to other jurisdictions. The e-ID issued to e-residents does not have the status of a passport or visa and does not automatically lead to Estonian permanent resident status, nor to citizenship in the traditional sense, although in a way it can be viewed as a new form of economic immigration. In establishing the program, Estonia has expanded its business and revenue base while keeping operational costs low. Estonia has opened new economic channels, created a new virtual domain for international commerce, and is fundamentally changing the nature of international commerce and finance.

2. A TRANSFORMATIVE SUCCESS STORY

The e-residency program has achieved its objective of economic expansion. The number of applicants for e-residency has grown steadily since the program launch, exceeding projections and expectations.[2]

There are currently 27,600 Estonian e-residents, who to date have established 4495 companies.[3] In a recent report, Deloitte estimated that e-residency has brought "€14.4 million in income, including €1.4 million in net income and €13 million in net indirect socio-economic benefits" to Estonia in three years.[4] This confirms the Estonian government analysis. The return on investment is estimated by the Estonian government to be €100 euros for each euro it has invested in the e-residency program.[5]

As has been the case since the inception of the program, Finland, Russia, Ukraine, and the U.S. have the largest number of Estonian e-residents. Overall, people from 138 countries have applied for e-residency.[6] There is now at least one Estonian e-resident in every corner of the world, making it one of the most expansive and comprehensive commercial networks in the world.

In the years following its launch, the e-residency program has developed rapidly in terms of number of applicants and the services available to them. New commercial services have been added through the Estonian government partnering with private-sector providers. Blockchain technology has been extended to identity authentication and verification, document authentication and management, payment systems, and new trading securities offered by NASDAQ, for example. Other developments include the establishment of eResNetwork, a new business networking platform for e-residents to communicate securely with other e-residents; and broader use of X-road, the platform used for both the e-residency program and e-Estonia services for physical residents and citizens, as a joint data exchange platform between Estonia and Finland.

---

1 Undistributed profits that are reinvested into the Estonian company are not subject to Estonian corporate tax.

2 Kaspar Korjus, the e-Residency Program Manager, reportedly stated in March 2017 that, "[It is] important for startups to set goals that are both ambitious and achievable. Our target of 10 million e-Residents will require exponential growth, but the early indications are that we are on schedule. We already have more e-residents than expected at this stage." As reported by Kalev Aasmae, "Estonia has 1.3 million people: Here's how it plans to get 10 million e-residents by 2025," ZDNeT, March 20, 2017, http://zd.net/2nQJWTa

3 Republic of Estonia, E-Residency statistics, http://bit.ly/1P68QaR.

4 Deloitte, 2017, E-residency brought €14.4 million to Estonia in first three years, December 2, http://bit.ly/2FVs3tL. According to this report, it is projected that by 2021, the program could generate €31 million in net income and €194 million in net indirect socio-economic benefits, assuming that Estonia will have 150,200 e-residents by 2021 who have established 20,200 businesses.

5 According to Kaspar Korjus, head of the Estonian e-residency program. See Deloitte (2017).

6 Republic of Estonia, e-residency, http://bit.ly/2BNJkqY

In 2018, Estonia also announced that it will be launching the first government-backed virtual currency, to be called Estcoin, to be used as part of the e-residency program. A number of approaches are under consideration. One approach is to use these "crypto tokens" to reward those who further develop the e-residency program, and refer new e-residents in furtherance of Estonia's objective of developing the digital nation. Another notable proposal is to use Estcoin for transactions by e-residents and perhaps others; and there are thoughts of pegging Estcoin to the Euro. Estcoin is seen as a means of reducing the costs involved in verifying identity for transactions and for avoiding cross-border banking fees for transactions among e-residents. The full implications are not yet known but it is clear that transactions using Estcoin will be blockchain-based and that this use of Estcoin will be a highly disruptive development likely to be also adopted by other nations.

3. BROADER SIGNIFICANCE

The relevance of the e-residency program extends beyond Estonia, as use of government-authenticated e-IDs for remote-access international transactions gains traction.

The Estonian program is setting the standard for similar international digital identity programs, most notably mutual e-ID recognition and data exchange between Estonia and Finland. Estonia, Belgium, Portugal, Lithuania, and Finland already mutually recognize their respective government-authenticated e-IDs for some transactions, and Estonia and Finland are further developing their interoperability using X-road, the exchange program used for e-residency, as well as e-services for Estonian citizens and physical residents.

The other major international development is the new "digital single market" (DSM) being established in the E.U. and the "single digital identity" (SDI) being established as part of that program. The objective of SDI is the mutual recognition of government-authenticated e-IDs between member nations to enable remote commercial transactions in the E.U. These initiatives are led by Andrus Ansip, a Vice President at the European Commission (E.C.), and the former Prime Minister of Estonia. Estonia's assumption of the E.U. Presidency in 2017 has further strengthened its influential role.

The Estonian e-residency program, its technology, and its commercial features are instructive for all nations and regions that want to expand their economic base without the security risks and costs associated with traditional immigration. In particular, the program offers many lessons for other nations in relation to new scope for economic expansion and development of international commerce and finance based on e-ID and blockchain technology. Many nations are considering the broader use of blockchain for identity management, for commercial and financial transactions, and for new cryptocurrencies.[7]

[7] Including, for example, the U.K., Australia, and the U.S.

By not defining itself by geography, but instead by digital capability, Estonia is re-writing what it is to be a nation in this era. By not defining economic participation primarily by physical location or birth, Estonia is changing traditional notions of residency and ultimately of immigration and citizenship, and is opening the way for universal e-IDs and global virtual economic citizens. In using blockchain as the basis of the e-residency program, Estonia is changing the way this technology has been used and is expanding its application to e-ID and a full range of commercial as well as financial transactions. In establishing new virtual services for e-residents, Estonia is substantially expanding commercial channels and is changing the nature of international commerce. In launching Estcoin Estonia is changing international finance.[8]

4. CHALLENGES

As the first international e-ID program, e-residency introduces new ways of doing business and new types of risk. Many of the features that make the Estonian e-residency program innovative and attractive to entrepreneurs also make it susceptible to misuse, especially for identity fraud, including the creation and use of new digital identities, transaction fraud, and trade-based money laundering. Estonia's start-up culture, which has had a notable impact on the development of the e-residency program from its inception, is of itself a risk factor. The Estonian government has candidly acknowledged that it is operating the e-residency program like a start-up and will address issues as, and when, they arise.[9]

The nature and magnitude of the risks largely depends on the accuracy, integrity, and security of e-program protocols and procedures, on its legal, regulatory and enforcement underpinning, and on the integrity of the program technology. A recent incident is illustrative. In December 2017, Estonia announced that it is upgrading the security of e-ID cards used by e-residents, as a result of a security vulnerability found in software installed on the embedded chip in cards issued between 16 October 2014 and 25 October 2017. The vulnerability affected cards and computer systems around the world that use these chips. The vulnerability in the chip on the e-resident e-ID made it possible for the e-ID to be misused, though Estonia has reported that it is not aware of any incidence where that occurred. The process for updating the certificates for the e-ID card as part of addressing the security vulnerability is proving to be slow, prompting Estonia to strongly recommend that in the meantime, e-residents use a private sector service called Smart-ID so that their business dealings can continue uninterrupted. Smart-ID is a mobile app that was launched in 2016 by SK ID Solutions, which is partnering with the Estonian government to issue certificates for identity documents held by e-residents, as well as Estonian citizens and physical residents. An e-resident can download the Smart-ID app to an Android or iOS phone and then only needs to authenticate his/her identity once using the e-ID card to access e-services.

Smart-ID can be used for transactions, signing agreements, and activating new cards for business banking and finance with LHV, an Estonian banking and financial services company; Swedbank a Nordic-Baltic banking group based in Stockholm, with a significant presence in Estonia; and with Leap IN, a business services provider that offers a turn-key solution for setting up a location-independent single-person company. While Smart-ID clearly facilitates business, it raises questions about the rigor of identity authentication and verification, especially for banking services that need to comply with the "know your customer" (KYC) and other monitoring and reporting obligations mandated by Anti-Money Laundering/ Counter-Terrorism Financing (AML/CTF) legislation enacted in most nations including Estonia following the 9/11 terrorist attacks in the U.S.[10]

Another move that raises similar concerns is a new one-step, one-time KYC process that will be used by Change Bank.[11] The bank will leverage the Estonian e-residency program's e-ID for identity authentication to quickly sign-up e-residents as customers for crypto-banking. Reportedly, this process requires only basic background information.[12] When an e-resident completes the simple one-step identity authentication, a multi-asset blockchain-based Change wallet is created, enabling use of cryptocurrencies by e-residents through a mobile app. A Change debit card allows e-residents to make payments and withdraw funds from ATMs all over

[8] See Republic of Estonia, "Estcoin: a proposal to launch the world's first government ICO," http://bit.ly/2EdsAHi.

[9] Siim Sikkut, ICT Advisor, Government of Estonia, "E-stonia – a startup country," Back Light, June 15, 2015 at http://bit.ly/2E8Tvnj

[10] The legislation generally mandates that banks and financial institutions check and report the identity of every customer. The KYC requirements demand that a person establish his/her identity to open a bank account usually through a face-to-face interview, at which time a birth certificate, passport, and other identity documentation is produced to authenticate and verify identity. The requirement for an initial face-to-face interview and subsequently for some specified transactions, is in line with the banks' obligations under Good Banking Practice, the Estonian banking code of practice, and with AML/CTF legislation. See Good Banking Practice, Part 6, http://bit.ly/2BLvTI0. Although this is not legislation, as a code of practice it closely follows the KYC and STR requirements typically found in the AML/CTF legislation.

the world, using a crypto-to-fiat currency conversion. It is envisaged that e-residents will eventually be able to use the Change mobile app to invest in stocks, obtain peer-to-peer loans, and buy and sell real estate.

The e-residency program's collaboration for these broader uses of blockchain also raises concerns, especially for the blockchain-based services that are capable of operating outside traditional legal frameworks and existing international monitoring and enforcement regimes. For example, Estonia is collaborating with Bitnation, one of several emerging initiatives based on blockchain technology that are specifically designed to bypass traditional, national governance systems. In its broadest application, Bitnation aims to use blockchain to provide a new system to vouch for identity, for contractual agreements, including those for banking and company incorporation, and for new payment systems that operate outside regulated, monitored channels.[13] As the joint press statement points out, "[v]ia the international Bitnation Public Notary, e-Residents, regardless of where they live or do business, will be able to notarize their marriages, birth certificates, business contracts, and much more on the blockchain."[14]

According to Susanne Templehof, founder of Bitnation, the broad objective is "to gain recognition for Bitnation as a sovereign entity, thus creating a precedent for open source protocol to be considered as sovereign jurisdictions."[15] This, in effect, seeks to "establish a new virtual jurisdiction with its own rules."[16] The underlying philosophy is that identity is established using a distributed ledger on a global open source platform, rather than using traditional authentication sources like government records and authentication intermediaries like banks. This potential use of blockchain for identity authentication and verification, and for at least some transactions for Estonian e-residents, is a significant development that can enable the provision of self-sovereign identity and other related services to e-residents, outside exiting legal channels and protocols. The development and use of these under-governed and ungoverned domains for commercial activity also increases their potential use for illicit, destabilizing activity that has both national and international implications.

E-residency introduces considerable change to international commerce that can, by its nature, be destabilizing. However, expansion of the program into global trading markets by adding NASDAQ and into new blockchain-based services offered by Bitnation are especially impactful, because anyone with an Estonian e-residency ID can engage in trade in stocks, futures, commodities, and currency. These developments have implications for the stability and security of global commercial and financial markets, and security generally. Of particular concern is the use of these types of programs for concealing and funding terrorist and organized criminal activity and other illicit and destabilizing activities by rogue individuals, foreign powers, extremist organizations, and criminal networks.

These aspects highlight the need for rigorous new security protocols and procedures; and for these types of international digital identity programs to be based within an effective, robust national and international regulatory and security framework designed to address the new challenges presented by these programs. This is particularly so in view of the program's international reach and impact.

---

[11] Change Bank is based in Singapore and is licensed by the Monetary Authority of Singapore. http://bit.ly/2nL6Fko.

[12] Prisco, G., 2017, "Estonia partners with Change Bank for blockchain e-residency program," NASDAQ, September 28, http://bit.ly/2GVMNDb.

[13] Bitnation describes itself as "a decentralized, open-source movement, powered by the Bitcoin blockchain 2.0 technology, in an attempt to foster a peer-to-peer voluntary governance system, rather than the current 'top-down', 'one-size-fits-all' model, restrained by the current nation-state-engineered geographical apartheid, where your quality of life is defined by where you were arbitrarily born." In further detail, Bitnation states that it "provides the same services traditional governments provides, from dispute resolution and insurance to security and much more – but in a geographically unbound, decentralized, and voluntary way. Bitnation is powered by Bitcoin 2.0 blockchain technology – a cryptographically secured public ledger distributed amongst all of its users. As we like to say – Bitnation: Blockchains, Not Borders." See Bitnation Governance 02 at http://bit.ly/2sesq0d.

[14] Tempelhof's comments in relation to the collaboration with Estonia are more moderate: "[m]y aim is to see a world where hundreds of thousands or millions of governance service providers in a free global market competing through offering better services at a better value, rather than through the use of force within arbitrary lines in the sand." "To that end, seeing nation state governments starting to provide governance services on a free global market as well, like The Republic of Estonia, is encouraging, and a step in the right direction. Now we need more nation state governments, as well as open source protocols joining the global market." See Giulio Prisco, "Estonian Government Partners with Bitnation to Offer Blockchain Notarization Services to e-Residents" Bitcoin Magazine, December 1, 2015, http://bit.ly/1OvfExl.

[15] As reported by Ian Allison, "Bitnation and Estonian government start spreading sovereign jurisdiction on the blockchain," 28 November, 2015, http://bit.ly/1OpDJpS. Bitnation has recently received international attention for providing assistance to Syrian refugees in Europe, including an emergency digital identity and financial services through a Bitcoin Visa card to enable a refugee who cannot establish a bank account to receive funds from family, for example. Blockchain is used to cryptographically establish an individual's existence and family relations to generate a digital identity. That identity generates a Quick Response Code, an optical label that contains information in machine-readable form that can be read by a mobile phone to apply for a Bitcoin Visa card, which can be used throughout Europe without a bank account. Susanne Templehof, founder of Bitnation, reportedly explained that "the Blockchain Emergency ID is a rudimentary emergency ID, based on the blockchain technology, for individuals who cannot obtain other documents of identification." She explains, "[w]e are providing emergency ID and then this visa card because most refugees will be unemployed. They won't be legally able to get a job for several years and they can't open a bank account." See Ian Allison, "Decentralised government project Bitnation offers refugees blockchain IDs and bitcoin debit cards" International Business Times, October 30, 2015, http://bit.ly/1RTRPR5. Use of blockchain in this type of situation to create an emergency, temporary digital identity to enable aid to be given to an individual who is unable to otherwise establish his/her identity may be admirable. However, it does raise security concerns, particularly in the use of this to create a new false identity and to engage in nefarious and covert activities ranging from crimes like money laundering to activities endangering national and international security.

[16] Ibid. As well as the huge increase in stateless people in Europe from the refugee crisis, Bitnation is looking at developing markets, assisted economies, and the grey economy. For example, the registry capabilities of blockchain are being considered as a means of recognizing land rights in the developing world, in countries like Ghana, where 70% of land is reportedly untitled and land is traded peer to peer.

A more comprehensive international approach is needed to establish a standard for the design and operation of these programs. Standards that currently exist are fragmented, tending to cover in depth either technical requirements,[17]or detailed procedural guidelines, such as those for banking, which are developed primarily at industry or sector level. As part of the DSM, the E.C is currently proposing a new E.U. Cybersecurity Agency and a certification framework to provide a set of rules, technical requirements, standards, and procedures. The focus, however, is on the E.U.[18]

The proposed certification is to attest that products and services are in certified in accordance with specified cybersecurity requirements, and will be recognized in all member states. This proposal is in its early stages so the time for implementation is not known, but it is designed to address areas of variability between member nations to facilitate trade across borders. It is also not yet clear how comprehensive or detailed the framework will be, but it is likely to be risk-based, in line with existing national and regional legal requirements that specify that there be appropriate technical and organizational measures to ensure a level of security appropriate to the risk to an individual's personal data. In line with the requirement in the E.U. to protect the fundamental human rights, the focus of the E.U. security requirements will primarily be on protection from unauthorized disclosure of the personal data, and privacy, of E.U. citizens and residents,[19] rather than on issues of concern to the broader international community.

Initiatives to address security issues are also E.U. focused. For example, the E.U. has recently adopted a framework for a joint E.U. diplomatic response to malicious cyber activities that sets out measures under the Common Foreign and Security Policy, including restrictive measures "that can be used to strengthen the E.U.'s response to activities that harm its political, security and economic interests."[20]

New transnational digital identity programs like Estonian e-residency and the E.U. DSM fundamentally challenge exiting regulation and enforcement, which is based on national law. The nature and broader effects of these programs require a coordinated international response. While these new e-ID programs raise new, more complex issues and concerns, the widespread adoption of the AML/CTF requirements is precedent for, and an example of, the type of international cooperation required and possible.

5. CONCLUSION

The next evolution of digital identity programs, like the Estonian e-residency and future iterations, have unprecedented implications for commerce, finance, security, international law, and legal norms, caused by the virtual dismantling of geographical boundaries and traditional concepts of immigration, residency, and even citizenship, based on birth and/or physical presence.

The technological, policy, and procedural vulnerabilities of this next evolution of digital identity programs impact both national and international security and stability. Identity fraud, fraudulent transactions, and money laundering, especially trade-based money laundering, are the most significant known risks. However, these types of digital identity programs and their data are strategic assets and potential targets that can be used by criminals and adversaries not just for known forms of identity fraud and money laundering, but for new, and as yet unanticipated, types of destabilizing and offensive activities. These programs have extensive international reach, with the Estonian e-residency program, for example, spanning the globe with potential operatives.

These new digital identity programs, of which Estonian e-residency is the current leader, offer many benefits to the host nation and to individual entrepreneurs, but they challenge the effectiveness of traditional regulation and, potentially, the stability and security of international financial and commercial markets. International coordination and cooperation is needed to ensure these programs meet international standards in their design and operation, particularly in relation to identity authentication at the time of program registration by an individual and, subsequently, for identity verification for transactions. International cooperation is also needed to establish a transnational framework for monitoring and regulating financial and commercial transactions, including those in currently under-regulated and unregulated channels and domains.

[17] The U.S. National Institute for Standards and Technology (NIST) technical guidelines for Federal agencies implementing electronic authentication is an example.

[18] See E.C., "Policies: digital single market: cybersecurity," http://bit.ly/2xAu7rq. See also E.C., "Digital single market, policies: the EU cybersecurity certification framework," http://bit.ly/2E9NOWj.

[19] For example, see Article 32 of the E.U. General Data Protection Regulation, which provides that "[t]aking into account the state of the art, the costs of implementation and the nature, scope, context and purposes of processing as well as the risk of varying likelihood and severity for the rights and freedoms of natural persons, the controller and the processor shall implement appropriate technical and organisational measures to ensure a level of security appropriate to the risk … account shall be taken in particular of the risks that are presented by processing … which could lead to physical, material or non-material damage" (my emphasis). Specific authoritative guidance is usually only provided in the event of litigation when a court may comment on requirements.

[20] The E.C. reports that "[i]mplementation work on the Framework is currently ongoing with Member States and would also be taken forward in close coordination with the Blueprint to respond to large scale cyber incidents." See E.C., "Policies: digital single market, cybersecurity, cyberdefense," http://bit.ly/2Eouubp.

The future of regulatory management: From static compliance reporting to dynamic interface capabilities

ÅKE FREIJ | Managing Principal, Capco

ABSTRACT

Historically, businesses have treated regulations as a necessary evil, and thereby managed them by reactive and siloed approaches towards minimum compliance. In this article, an approach for "future regulatory management" is presented. From an overview of how regulations have evolved over time, an analytical framework is applied to outline the capabilities required for managing regulatory change in the future. In addition, we offer five design principles that will give firms a chance to innovate with regulatory change rather than just continue to fight with compliance requirements. Instead of being viewed as "the perpetual ogre, the bad guy who is against good things" [Levitt (1968)], you could be the company that customers turn to as a role model.

1. THE FUTURE DEMANDS NEW CAPABILITIES TO MANAGE EVOLVING REGULATIONS

Since the financial crisis of 2008, discussions about regulations have focused on the increasing burden and the difficulties companies face in remaining compliant [Arner et al. (2016), Gerlach et al. (2016)]. There is nothing new in that perspective. Businesses have always treated regulations as something that needs to be avoided or minimized [Levitt (1968)]. Flagrant examples of avoidance include food poisoning, workplace safety, and even child labor [Minzberg (1984)]. Instead of viewing regulations in a positive light and being proactive, companies tend to implement regulations in silos, be reactive, use checklists, and apply point technology solutions [Freij (2017)]. Some financial institutions continue to evade regulations through the use of the so-called "shadow banking" [Worstall (2015)], spending resources to avoid rather than embrace and implement new regulatory requirements. To change the approach to managing regulations, companies need to look beyond executing strategic political management [Oliver and Holzinger (2008)], engaging in regulatory avoidance [Fox-Wolfgramm et al. (1998)], and influencing authorities via regulatory capture [Dal Bó (2006)]. A new set of guiding principles for building capabilities is needed. Instead of a burden, regulatory change can be seen as a trigger for radical innovation [Bieck and Freij (2010)], creating opportunities for "first mover advantages" and innovation [Lopez and Roberts (2002)].

This article will look back at the history of financial regulations and attempt to apply innovation theory to understand its implications [Fagerberg et al. (2012)]. Innovation is rarely considered by managers (and researchers) when discussing regulations [Frame and White (2004)]. By using innovation as a starting point, firms can manage the implications of regulations in a strategic manner to generate value and advantages vis-à-vis competitors. In order to direct such efforts, six capabilities are outlined to support the strategic and operational management of the impact of regulatory change. These capabilities form the basis for creating a strategic regulatory management function. By acquiring, nurturing, and executing these capabilities, firms and regulators can view the financial services industry not as internal silos based on single products and processes, but instead as a dynamic ecosystem with interconnected institutions [Jacobides et al. (2014)].

Understanding how regulations evolve, and how companies address them in different ways, will help both regulators and managers make better decisions in the future, hence avoiding repetitions of previous crises [Jacobides and Winter (2010)].

2. AN ANALYTICAL FRAMEWORK TO UNDERSTAND THE IMPACT OF REGULATIONS

The events depicted in research on regulatory change underscore the importance of viewing regulations as a source of change and the need to understand their impact, as well as any requisite response. The complex dynamics involved when regulations change mean that they need to be viewed and managed as new tasks. Firms are exposed to a variety of challenges as they move from grasping the impact of a regulatory change on operations to the corresponding implementation of the requirements. The complications associated with implementation are due to the fact that "changing regulatory requirements are creating a derived, albeit uncertain, demand" [Pilkington and Dyerson (2004)].

2.1 Firms face difficulties in implementing regulatory change

The complex implications of a change in regulation make it difficult for firms to manage the implementation of the corresponding requirements. Regulatory change creates different types of new requirements [Abernathy and Clark (1985)], influences the role of new and existing products and services, as well as how they are connected [Henderson and Clark (1990)], and results in new processes that affect the role of internal and external providers and the interfaces between them [Jacobides and Winter (2005)].

The ways in which regulatory change influences firms has been observed in various industries. In radio broadcasting, regulatory changes gave new firms a chance to enter the market by exploiting new products, processes, and technology [Funk (2015)]. In the mobile internet market, regulations have influenced how firms introduce new services into the market [Tee and Gawer (2009)]. Similarly, regulatory change in the financial services industry has led to launching new products and processes [Jacobides (2005)].

2.2 Influence of regulations on management tasks

Research on regulatory change, and its implementation, point to three management tasks that describe the challenges firms face. The first task is to understand the industry dynamics. This task describes how the logic of an industry can be affected by regulatory change. The second task is to consider the relative position of firms [Funk (2015)]. The change in industry logic provides options for firms to find new roles in the value chain. The third task relates to integration of operations, in the form of arrangements within and between firms [Jacobides (2005)]. Examples of implications in this category include new products and processes, new sourcing arrangements, and new forms of collaboration.

2.3 Areas of actions when regulations change

Actions in relation to regulations and regulatory change are examined in studies investigating how the operations of firms are impacted. The implications can alter the focus of attention of a firm or business operations [Teece (1986)] by introducing significant modifications to products, processes, and technology (Figure 1). The evolution of regulations can prevent firms from implementing products and services as intended [Penrose (1959)].

Regulatory change impacts operations because of the associated implementation of new requirements. Impact has previously been noted in such areas as product development [Brown and Eisenhardt (1995)], customer and user requirements [Oliveira and von Hippel (2011)], and evolution in technology [Anderson and Tushman (1990)]. Internal research and development activities can look to regulations and changes in regulations for guidance and evaluation of new solutions [Nelson and Winter (1982)]. The role of customers and users in the market is also modified when regulations change, since their requirements might be updated, and firms translate these requirements and integrate them into product and process offerings [Richard and Devinney (2005)].

Table 1: An analytical framework to understand the impact of regulations

	PRODUCT DESIGN	PROCESS ORCHESTRATION	TECHNOLOGY PLATFORMS
Industry dynamics and logic			
Firm role in value chain			
Integration in operations			

3. EVOLUTION OF FINANCIAL SERVICES REGULATIONS

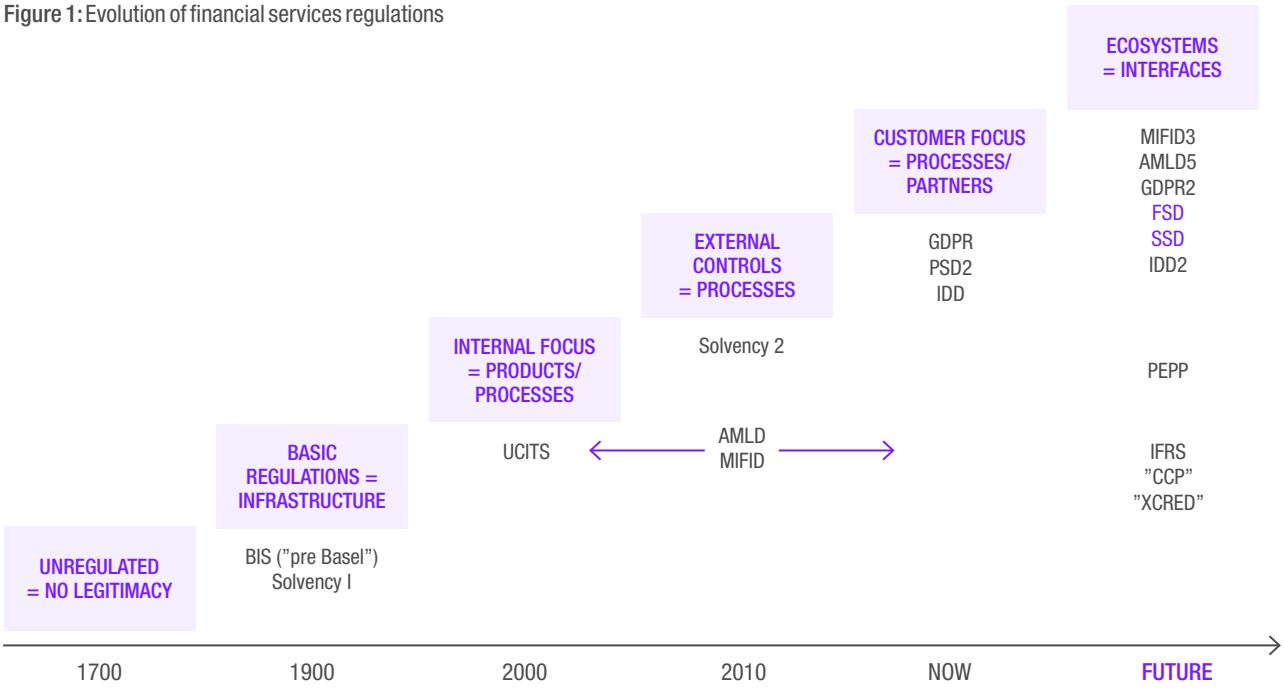
3.1 Current state of the debate on financial services regulations

Financial services executives have in recent years been complaining that the industry is “under assault by regulators” [Son (2015)]. This is an industry in which regulations and regulators frequently play a significant role in the evolution of firms and how they manage their business.

How firms view regulatory change in the financial services industry could be compared to “watching an arms race, a contest in which the rules get ever-more complicated as well-resourced banks try to outflank regulators and regulators try to catch up” [Wessel (2012)]. As this article is written, a typical financial institution is dealing with around 40 different regulatory changes [Moreno (2014)]. Most of these regulations are implemented at a central level in the firms, as well as by each business unit and local subsidiary. The complexity of the combined regulatory pressure could lead to the existence of up to a thousand different projects in each firm, where the potential benefits might reside in the individual project or in the combination of steps to implement two or more regulatory changes.

It should be added that while most of the public ire has been focused on banks, major insurance companies have also needed to be bailed out, such as AIG in the U.S. [Harrington (2009), Klein (2012)]. Due to these events, the insurance industry is subject to a growing list of regulatory changes as well.

Figure 1: Evolution of financial services regulations



3.2 Historic evolution of the financial services industry (and its regulations)

It can be argued that the first modern financial services organizations emerged as a result of the industrial revolution in the 18th century. The introduction of salaried labor drove their establishments so that available money could be deposited and withdrawn. At this time, institutions such as "widow and orphan funds," mutual fire insurance associations, and collective savings banks were largely unregulated [Lindmark et al. (2006)].

With financial firms growing larger, basic regulations to assure a fundamental infrastructure were introduced throughout the early part of the 20th century. This included the establishment of the Bank for International Settlements, as well as the introduction of Solvency 1 for insurance firms. The question asked here was: "Do you have the basic funds and structure required to support your business?" Firms needed to report balance sheets, income statements, capital reports, and cash flow analyses.

In another phase, moving into the 21st century, focus moved to better understanding of internal products provided and processes performed. Disclosures were demanded to cover more than the pure financial risk, and started to cover larger parts of internal operations. The introduction of regulations such as UCITS [De Smet (2012)] and MIFID1 contributed to an increased understanding of what firms offered customers and how the offer was presented.

In order to further shed light on how financial firms acted, the next area of focus was an extension of regulations towards external processes. This requirement came in the shape of regulations such as AMLD4 and MIFID2, where not only internal activities were scrutinized but also links to parties such as agents, intermediaries, and network connections were assessed.

The most recent evolution of regulations introduces a core of customer focus. A common theme is articulated as "consumer protection." Regulations such as the General Data Protection Regulation (GDPR), the second Payment Services Directive, and the Insurance Distribution Directive all emphasize the importance of protecting the customer against misuse of data and profit maximization from the financial services providers. This also includes increased scrutiny of the role of salespeople and advisors.

The coming era of financial services regulations will introduce further complexity through increased focus on ecosystems and interfaces. We might see a coming generation of regulations extending the demand for control across firm boundaries (could appear as, for

example, MIFID3, GDPR2, AMLD5) and further "open financial services" regulations (could be variants of PSD2 for funds and insurance in the form of Pan European Pension Products (PEPP) – opening for full flexibility and account transfers across the E.U.). On top of this added complexity, we will see additional evolution of basic reporting requirements (examples include new issues of IFRS, increasing counter-party reporting, and detailed reporting of credits given, like Anacredit). Finally, there are complex themes emerging on regulating fintechs [Alvarez (2017)], crowdfunding, sustainable investments/fiduciary duty, and the entire "digital economy" initiative from the E.U. (covering big data and internet of things).

3.3 Regulation, de-regulation, or re-regulation?

Will regulations for financial services increase or decrease (or even stay the same)? Many are hopeful that deregulation is on the cards and will lower the burden on the industry [Paletta (2017)].

It should be said, however, that the term "deregulation" is slightly misleading, since the removal of a regulation usually involves its replacement by another that may be perceived as allowing more (but for some actors less) innovative activities. Further, to characterize an industry as "regulated" or "unregulated" [Wiseman and Catanach (1997)] might be an oversimplification, since no industry is devoid of regulation. The level of regulation per se is of less importance than the implications of regulatory change. Firms will have difficulty forecasting the direction of change [Grasshof (2015)], hence preparation for the change is what will make firms able to consider novel actions.

As outlined above, the evolution of regulations over time (from unregulated to a focus on ecosystems and interfaces) has gradually evolved from internal implementation issues with a major focus on technology to a more complex issue of industry dynamics across functions in the entire company. As such, the different eras of regulatory development can be related to the analytical framework presented earlier. This historical picture forms the basis for identify six overarching capabilities for regulatory management.

Figure 2: Evolution of regulations in the analytical framework

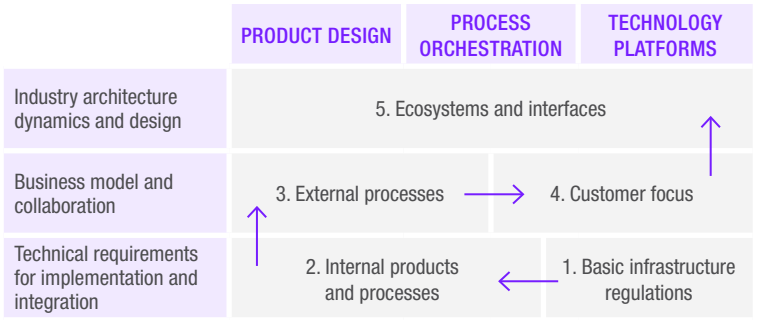
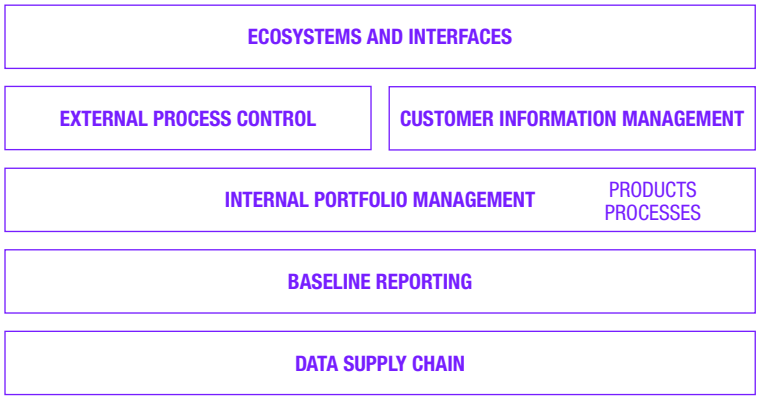


Figure 3: Capabilities required to capture value from regulatory change



4. CAPABILITIES REQUIRED TO CAPTURE VALUE FROM REGULATORY CHANGE

Based on the theoretical foundation, the analytical framework, and the observed historical evolution, six capabilities are identified to support the management of regulations and regulatory change in the future. The capabilities range from providing baseline reporting to the management of ecosystems and interfaces. Each capability contains a number of underlying functions, and examples of those are given below.

4.1 Data supply chain

As a foundation for future regulatory management, and any change that will impact the business, a platform to achieve commonalities across different business units and regulatory requirements is valuable. Financial services companies can be seen as a type of information processors and data handlers [Jacobides and Winter (2005)]. If the basic manufacturing facility for data and information is not in place, it will be difficult to deliver high quality products and services to customers and partners.

Underlying functions to the data supply chain includes design of standardized data feeds, the use of artificial intelligence and machine learning, ETL (Extract, Transform, Load) processes for data, development of data lake(s), and an underlying and supporting ontology (including data models and glossaries).

4.2 Baseline reporting

The fundamental reporting of the status of a business is the essence of a modern functioning capitalist society [Baldwin (2008)]. A financial services firm is subject to additional scrutiny due to the responsibility to manage "other people's money." Increasing levels of oversight have evolved across the introduction of Basel 1, 2, and 3 [Jones (2000)] as well as the corresponding insurance regulations Solvency 1 and 2 [Klein (2012)]. Recent developments have been triggered by the requirements from EMIR in securities processing.[1]

Baseline reporting includes functions to report on topics such as financial statements, risk management, credit data, and counterparties.

4.3 Internal portfolio management

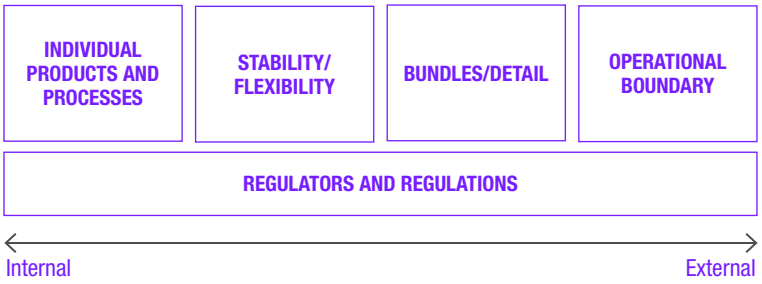
A portfolio with a range of internal processes and products needs to be managed as part of several regulations. This is applicable for fund products under UCITS, PRIIP, and also numerous local product centric regulations. Processes are central to earlier versions of AMLD and MIFID and parts of the Solvency 2 and Basel 2 and 3 accords.

For products, requirements should support product approval, performance management, and component sourcing. Process capabilities can include robotic process automation, process optimization, and performance management.

4.4 External process control

The increasing importance of regulations not only covering the internal company scope, but also looking at the process beyond the boundary of the firm, increases the need for standardization [Hülsse and Kerwer (2007)]. The complexity and severity of not managing this context is increasing [McIntosh (2016)]. Regulations, such as AMLD4, MIFID2, as well as sections of Basel 3 and Solvency 2, contribute to the change in scope.

Figure 4: The capability to manage interfaces



Quality assurance, fraud management, and business process optimization are examples of functions needed for this capability.

4.5 Customer information management

Taking the next step in evolving regulations, the need to have capabilities to support customer information management should be considered. The functions here are not the same as those needed in a market oriented CRM capability (but there are certainly overlaps and synergies). Functions needed are crowdsourcing, 360 view (both from the inside and the outside), and customer journey management.

4.6 Ecosystems and interfaces

The future of regulatory change in a coming generation is expanding into the need to manage ecosystems and interfaces. The prediction is that further regulations will appear in what follows PSD2 and GDPR. In addition, continued dynamics in flow across country borders will be seen (as an example the emerging discussion about pan-European pensions, PEPP, can be put forward).

Functions required in this capability are: API management, business model innovation, alliance management, and co-creation.

The journey over time that resulted in defining the capabilities needed for the future management of regulations point to the need to understand and manage interfaces. An excursion is, therefore, made into how a capability to manage interfaces could be described.

[1] TradeChannel, 2017, "New RTS and ITS for EMIR published," January 25, http://bit.ly/2tx1IGz

4.6.1 THE CAPABILITY TO MANAGE INTERFACES[2]

The aforementioned six capabilities are outlined for successful management of historic, current, and future regulations and regulatory change. The inherent functional requirements inherent span from individual and internal products and processes to the management of the operational boundaries of the firm. The nature of these capabilities is in turn dependent on the design of a capability to manage interfaces.

When the new requirements from a regulatory change are implemented, the implementation serves as a new basis for addressing the next round of regulatory changes. In the current business environment (certainly in financial services, but also in other sectors such as the automotive and transportation), the queue of regulatory changes is mounting, so it is likely that a constant flow of regulatory changes will need to be managed.

A challenge in managing interfaces is that they are generally invisible, functioning as links between the (more visible) interacting parts of a system that they support. The detection of interfaces emerging because of increased integration from the impact of regulatory change is a step towards treating the interfaces as just as visible as the parts (products, processes, and technology) that were connected. In the course of establishing a design (which could be a new product, process, or technology), "the detailed interface specifications... need to be set in advance and known to the affected parties;" hence it is important that "interfaces are visible information" [Baldwin and Clark (2000)]. Interfaces describe in detail how functions in a system interact, including how they will fit together, connect, and communicate [Baldwin and Clark (1997)]. Interfaces are thereby required for integration to be established on different levels.

4.6.2 INTERFACES WITHIN INDIVIDUAL PRODUCTS AND PROCESSES

The initial action taken by firms after a regulatory change is to ensure that the basic compliance requirements are met. New features are added to products based on the content of the regulation. Processes will require new steps to fulfill the requirements from the changed regulation, or alternatively new processes may be implemented. Integration and corresponding interfaces within individual products and processes developed may include links between different product components and individual tasks in the new process. The function of the interfaces in an individual process is to facilitate handovers across different units or departments involved in the process. Attention to interfaces even within individual products and processes is needed, since the requirements arising from the regulatory change can be of a different nature from what the firm has been accustomed to managing before the change.

[2] This section is based on a more detailed outline in Freij (2017).

4.6.3 THE BALANCE BETWEEN STABILITY AND FLEXIBILITY REQUIRES INTERFACES

Following the initial actions taken to meet the requirements within individual products and processes, increased emphasis is placed on integration between the new products and processes and the existing ones. When the combination of new and existing products and processes demands flexibility, but also needs to maintain the previous stability, an increased focus on the impact on associated technology interfaces is required. In this stage, additional actions can be taken concerning product management to balance flexibility and stability in both new and existing products. The exposure of new functionality in products introduced in response to a regulatory change creates a risk that customers will lack understanding of the new offerings. Actions are taken here to determine to what extent the available products from before and after the regulatory change should be presented as joint offers to the customers in the market.

4.6.4 INTERFACES TO INTEGRATE BUNDLES AND DETAILS

A regulatory change introduces the need for products, processes, and technology to be broken down in more detail or, alternatively, allows options that are more aggregated than before the change. One approach to deal with aggregating detailed parts of a solution is bundling, where the firm decides which combinations of products, processes, and technology to provide. The level of balancing between bundles and details depends on the requirements in the regulation. The development of interfaces in the evolution towards increased integration after a regulatory change is necessary to maintain a balance between bundled solutions and the introduction of products and processes that are broken down and presented to customers in more detail. When an existing process for a bundled offer towards customers is integrated with a new process that (conversely) breaks down customers' options into more details, interfaces to integrate the two different processes are facilitated to manage the implementation of new requirements.

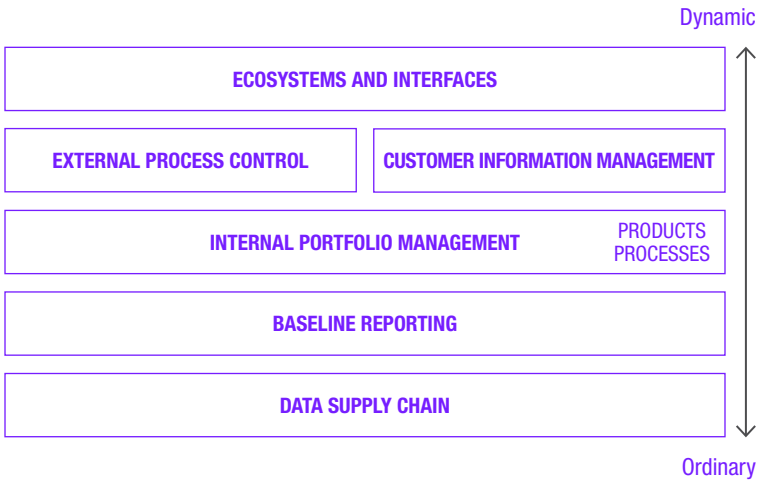
4.6.5 OPERATIONAL BOUNDARY INTERFACES

Towards the end of the evolution after a regulatory change, the focus turns to interfaces that address the operational boundary between the firm and adjacent actors. As new products, processes, and technology are integrated with existing offerings, customer involvement increases since the customer has more options to choose from. Customers are also more involved in the decisions related to the configuration of the offering based on products and processes so as to provide more flexibility as a result of the regulatory change. The need to integrate customer requirements in the products that previously have been managed only internally gives rise to new sequences of tasks (and also new tasks). Furthermore, new information is needed to match the customer's functionality requirements to the products provided. In addition, customers demand information about their specific situation and how it relates to their engagement with the firm and its products. Hence, it is not sufficient to communicate the same general product information for all customers. The increased focus on customers is due to the options available for them to make selections within the products available, which has associated repercussions for the management of product support processes and distribution. Tasks that match customer requirements may be performed with a higher or lower frequency (e.g., daily instead of yearly, or vice versa) as a result of a regulatory change.

4.6.6 INTERFACES TO UNDERSTAND REGULATORS AND REGULATIONS

After a regulatory change occurs, firms attempt to understand its implications and define their approach to implementing the new requirements. Each firm needs to understand relationships to other current regulations and what parts of the organization are influenced to determine its approach to implementation. In addition to considering the regulatory change in itself, the forces behind it also form part of the understanding, since

Figure 5: Capabilities to manage regulations range from ordinary to dynamic



differences in the process leading up to the regulatory change influence the actions taken by individual firms. These forces include lobbying, political desires, deregulation interests, and customer requirements. Competitors could act to infuse requirements related to specific products, processes, and technology into the regulatory change. If one's own firm has been involved in the activities leading up to the regulatory change, such activities form part of the understanding of the context of the change. Also, the views of political actors and customers will be reviewed to grasp the potential influence of the change.

4.7 An intricate balance of regulatory change: ordinary and dynamic capabilities

A regulatory change has two contradicting implications for firms, in that it can both create restrictions and open up new opportunities for changing the position of the firm. As has been noted, firms that manage to deal with the impact of regulatory change are in possession of the capability to manage interfaces.

Capabilities have previously been categorized into two types, ordinary and dynamic [Teece (2014)]. An ordinary capability is the basis for performing administrative tasks, such as compliance with regulations. A dynamic capability is applied to manoeuvre in a changing business environment and to orchestrate resources. The capability to manage interfaces as applied by successful firms after a regulatory change spans a range from ordinary to dynamic, which presents a difficulty for the management of new requirements. Regulations demand administrative capabilities to comply in an on-going operation, but when regulations change, the required capability shifts towards an entrepreneurial emphasis due to the intricate influences presented [Penrose (1959)]. The capability to manage interfaces is thereby related to the possession of institutional assets needed to manage the relationships with regulations and regulators.

The capability to manage interfaces is, therefore, both ordinary (administrative) and dynamic (entrepreneurial). Firms in possession of the capability to manage interfaces are better prepared to manage shifts in focus from pure compliance to understanding the impact on new products, processes and technology.

Moving forward, when summarizing the historical impact (and potential future requirements) of regulations and the capabilities and functions articulated above, five design principles are defined for the future of regulatory management.

"There is an opportunity for financial services actors to harvest more value from regulatory implementations by applying a proactive, holistic, and constructive approach."

5. FIVE DESIGN PRINCIPLES FOR REGULATORY MANAGEMENT OF THE FUTIRE

There is an opportunity for financial services actors to harvest more value from regulatory implementations by applying a proactive, holistic, and constructive approach. When reflecting the historic development and the current initiatives in the pipeline, five design principles are identified to guide improved management of regulations in the future. The principles are: 1) digital economy grade data quality, 2) instant counterparty management, 3) full API connectivity, 4) customer interaction leverage, and 5) regtech plug-in architecture. The principles need to be addressed in their own right, but also constitute the building blocks of a strategic regulatory management function for the company.

5.1 Digital economy grade data quality

The maintenance and evolution of data to support the increased demands of baseline reporting will continue and will also be extended. The introduction of new accounting and reporting standards through IFRS (including IFRS9 for financial instruments and IFRS17 for insurance contracts) is one theme in this domain. Another is the further need for reporting on detailed credit positions (such as AnaCredit). In addition, the requirements for keeping reporting constant and flexible is increasing (see, for example, the Capital Requirements Directive no 4). It is both a matter of showing the numbers in static form, as well as providing

dynamic solutions for regulators to undertake their own analyses. One tangible example of this development is found in Austria [Wolf and Huber (2017)]. In addition, quality of the data needs to improve, where examples have demonstrated the need for improvement (e.g., in Sweden, where the supervisor "identified extensive quality deficiencies in the reporting of insurance firms").[3] Further, firms need to provide data lineage to show where the end result of the data reported comes from. This requirement is similar to when textile manufacturers needs to be aware of where their goods are produced (and if, for example, child labor is used in their production).

The increased focus of processing data and information is driven by, and inherent in, regulations such as BCBS239, GL44, Basel 3, Solvency 2, and GDPR. This drives the need to establish the capability for a data supply chain platform. Advancing data quality towards "digital economy grade" will be demanded to meet requirements of current (GDPR, PSD2, AMLD4, MIFID2, and EMIR) regulations better. In the future, for emerging regulations (such as, ePrivacy, other API driven regulations, AMLD5 – 6, and a potential revision towards MIFID3) the heightened threshold for data quality is simply a must-do.

5.2 Instant counterparty management

Over time, the gradual need to manage relationships with (and report on the status of) counterparty arrangements has grown. Implications of not having such arrangements in place were very visible in the wake of the 2008 financial crisis [Harrington (2009)]. Since then, the complexity of the "everyone-to-everyone" economy, with business and financial relationships has accelerated further. The increase in requirement in reporting data about counterparty exposure is, therefore, a second design principle for the future of regulatory management. The requirements are visible in, for example, Central Counterparty provisions under EMIR. The building of platforms for technical security and reliability can support the need for managing systemic risks in the financial services industry [Lopez (2017)]. The requirements are also driven by reduced cycle times in financial transactions [Abel (2016)] and the emergence of increased transparency in the ecosystem [Lenz (2016)]. Increased demands for understanding who are involved as counterparties in your business are in sight as a result of ESG (environment, society, governance) and the related sustainable investment policies [Rust (2017)]. Hence, solutions and systems need to be designed with the goal of always being able to determine the actors involved in any transaction.

5.3 Full API connectivity

Regulators and consumers have started to see the need for more transparency in the financial services industry. In 2018, a first step is taken for financial services firms to become significantly more open with regards to client data. This drive relates to both product and transaction data as well as personal data. Products and transactions are getting more open by the introduction of PSD2, which prescribes requirements to give other actors in the ecosystem data about what your customers have been doing. The management of personal data is entering a new phase with the launch of GDPR.

In the aftermath of implementing both PSD2 and GDPR in 2018 there will be two issues to consider. The first is operational, where a need to handle incidents and claims will surface. An assumption is made here that there is a lack of preparation with regards to the requirements outlined, and that there is significant scope for interpretation of regulatory requirements. The second is strategic, and relates to the need for the management of interfaces and ecosystems to be elevated. New functions and tools for the evolution of business arrangements need to be put in place [Jessel (2016)].

When APIs are increasingly published and connected, new forms of business can be executed [Egner (2017)]. Here banks and other financial institutions could engage in creative business models. One example of such models is to establish a position as "quality controller / certification authority" of data and relationships.

5.4 Customer interaction leverage

A mantra in regulatory and risk circles in recent years has been "know your customer" (KYC). In essence, the notion of knowing your customer is positive, and should lead to dedication and energy from financial services firms. Currently, on the contrary, the view of "KYC" is a burden of populating forms and asking numerous questions from the client. In the future, when continued operationalization and integration of processes related to MIFID2 and AMLD4 evolve, the data captured about the customer will need to be "collected once, used multiple times." This design principle further enlightens the need to better

[3] Finansinspektionen, 2017, "Reporting of insurance firms (summary)," November 7, http://bit.ly/2FCYwbQ

orchestrate processes connected to the relationships with customers. Included here is the ability to combine insight that is needed for requirements to report on, for example, money laundering with offering of better solutions to customers in terms of products and services. An important task here will be to better align regulatory compliance investments with customer relationship management initiatives. A driver is the increasing amount (and complexity) of big data, where needs are imminent to better make sense of massive data sources [Sidhwani (2016)]. Customer interaction leverage implies that data should be captured only once and then published across products, processes, platforms, and interaction points. This can significantly improve the depth and quality of financial advice, and allow for an efficient coexistence of supporting technology and human-centric relationships [Davenport (2017)].

5.5 Regtech plug-in architecture

To support the requirements to better manage the impact of regulations and regulatory change, "regtech," which is emerging on a larger scale as a spin-off from the fintech scene [Larsen and Gilani (2017)], has evolved. The growth in regtech solutions, providers, and technologies show indications for promises but also challenges [Weber (2017)]. Here, both incumbent technology actors, as well as new entrant entrepreneurs, offer solutions for one or more of the capabilities outlined above. Technologies are both established (such as analytics and process automation) but also emerging (such as artificial intelligence, machine learning, and blockchain) [Treleaven and Batrinca (2017)]. An imperative for building future capabilities will be the utilization of potential solutions offered by the regtech ecosystem. Financial services firms seeking to benefit from regtech solutions should develop an architectural vision and plan for which capabilities need support and how to integrate regtech components in their current and future solution architecture. A regtech buyer should also consider potential generic reuse and not just buy for point compliance. In order to do this, an approach based on functional match, architectural alignment, and reuse is fundamental to success [Butler (2017)]. Hence, we can avoid duplication in capability delivery and technical functionality. Each firm collaborating with regtechs should have a clear view of where the solution should be plugged in, from both process and functional / technical architecture perspectives.

CONCLUSIONS AND RECOMMENDATIONS

Changes in regulations, and associated requirements, will continue to play a significant role in driving innovation, changing the industry structure, and forming a large part in the investment budget of financial services firms. What can firms do to get more leverage from their investment, and also promote their business positions? Three actions are proposed here: 1) establish a function for strategic regulatory management, 2) consider innovation effects in every regulatory project, and 3) follow the five regulatory management design principles identified.

To establish and operate a function for strategic management of regulations, the focus should be on acquiring the six capabilities outlined earlier. To understand how to develop these capabilities you first need to investigate the historic approach taken by your company in relation to implementing regulatory requirements. You can refer to the historical evolution outlined in this article. In this step, you should also consider the business strategy, business model, and priorities of the company. Secondly, an analysis of a selected number of past, current, and future regulations, relative to your status under each capability should be performed. Thirdly, the identification of gaps to bridge (and the value of investing in them) needs to be performed. In the fourth step, you investigate and discover the "regtech" ecosystem to identify potential partners and solutions by considering capabilities supported, regulatory coverage, and enabling technology. After these four steps are performed, you can create a roadmap, an action plan, as well as a list of quick-wins to be realized on the road towards "regulatory management of the future."

Organizations that adopt the ideas outlined in this paper will benefit by making more efficient investments into meeting regulatory requirements. In addition, they will find directions for capturing value in the fast-changing industry landscape. Finally, they will see clearly that there is no contradiction between innovation/digitization projects and compliance projects. Instead, the realization will grow that the two seemingly contradicting investment streams support the same goal in the long run – satisfied and profitable customers. The framework presented here can guide both firms and regulators to better understand the actual effects of technological innovation and the real effectiveness of financial services regulations [Kane (1981)]. In the end, the hope is that society stops looking at financial services firms as "the perpetual ogre, the bad guy who is against good things" [Levitt (1968).]

References

Abel, J., 2016, "Time is risk: shortening the U.S. trade settlement cycle," Journal of Financial Transformation 44, 8-12

Abernathy, W. J., and K. Clark, 1985, "Innovation: mapping the winds of creative destruction," Research Policy 14, 3–22

Alvarez, C., 2017, "Fintech regulation: seeking a level playing field for everyone," BBVA, November 22, http://bbva.info/2leDu1E

Anderson, P., and M. L. Tushman, 1990, "Technological discontinuities and dominant designs: a cyclical model of technological change," Administrative Science Quarterly 35, 604–633

Arner, D. W. J., J. Barberis, and R. P. Buckley, 2016, "The emergence of Regtech 2.0: From know your customer to know your data," Journal of Financial Transformation 44, 79-86

Baldwin, C. Y., 2008, "Where do transactions come from? Modularity, transactions, and the boundaries of firms," Industrial and Corporate Change 17:1, 155–195

Baldwin, C. Y., and K. B. Clark, 1997, "Managing in an age of modularity," Harvard Business Review 75:5, 84-93

Baldwin, C. Y., and K. B. Clark, 2000, Design rules: the power of modularity, MIT Press

Bieck, C., and A. Freij, 2010, "Solving the innovation puzzle," IBM Institute for Business Value, https://ibm.co/2oY6OBr

Brown, S. L., and K. M. Eisenhardt, 1995, "Product development: past research, present findings, and future directions," Academy of Management Review 20:2, 343–378

Butler, T., 2017, "Towards a standards-based technology architecture for regtech," Journal of Financial Transformation 45, 49-59

Dal Bó, E., 2006, "Regulatory capture: a review," Oxford Review of Economic Policy 22:2, 203–225

Davenport, T., 2017, "To robo or not to robo: The rise of automated financial advice," Journal of Financial Transformation 46, 46-53

De Smet, D., 2012, "Exploring the influence of regulation on the innovation process," International Journal of Entrepreneurship and Innovation Management 16:1-2, 73–97

Egner, T., 2017, "Open APIs and open banking: assessing the impact on the European payments industry and seizing the opportunities," Journal of Financial Transformation 45, 8-13

Fagerberg, J., M. Fosaas, and K. Sapprasert, 2012, "Innovation: exploring the knowledge base," Research Policy 41:7, 1132–1153

Fox-Wolfgramm, S. J., K. B. Boal, and J. G. Hunt, 1998, "Organizational adaptation to institutional change: a comparative study of first-order change in prospector and defender banks," Administrative Science Quarterly 43:1, 87–126

Frame, W. S., and L. J. White, 2004, "Empirical studies of financial innovation: lots of talk, little action?" Journal of Economic Literature 42:1, 116-144

Freij, A., 2017, "Mastering the impact of regulatory change: the capability of financial services firms to manage interfaces," PhD Thesis, Stockholm School of Economics

Funk, J., 2015, "Industry architecture, the product life cycle, and entrepreneurial opportunities: the case of the US broadcasting sector," Industrial and Corporate Change 24:1, 65-91

Gerlach, C.A., R. Simmons, and S. Lam, 2016, "US Regulation of FinTech-Recent Developments and Challenges," Journal of Financial Transformation 44, 87-96

Grasshof, G., 2015, "What's the future of global banking regulation?" World Economic Forum, May 5, http://bit.ly/2oVymYs

Gurses, K., and P. Ozcan, 2014, "Entrepreneurship in regulated markets: framing contests and collective action to introduce pay TV in the US," Academy of Management Journal 58:6, 1709-1739

Harrington, S. E., 2009, "The financial crisis, systemic risk, and the future of insurance regulation," Journal of Risk and Insurance 76:4, 785–819

Henderson, R., and K. Clark, 1990, "Architectural innovation," Administrative Science Quarterly 35, 9-30

Hülsse, R., and D. Kerwer, 2007, "Global standards in Action: Insights from Anti-Money Laundering Regulation," Organization 14, 625-642

Jacobides, M., 2005, "Industry change through vertical integration," Academy of Management Journal 48:3, 465-498

Jacobides, M., and S. Winter, 2005, "The co-evolution of capabilities and transaction costs," Strategic Management Journal 26:5, 396-413

Jacobides, M. G., and S. Winter, 2010, "Survival of the reckless: how the US mortgage market evolved toward disaster," Paper presented at the Academy of Management annual meeting, Montreal, Canada

Jacobides, M., M. Drexler and J. Rico, 2014. "Rethinking the future of financial services: a structural and evolutionary perspective on regulation," Journal of Financial Perspectives 2, 1-26

Jessel, B., 2016, "The rise of the interconnected digital bank," Journal of Financial Transformation 44, 67-78

Jones, D., 2000, "Emerging problems with the Basel Capital Accord: regulatory capital arbitrage and related issues," Journal of Banking and Finance 24, 35-58

Kane, E. J., 1981, "Accelerating inflation, technological innovation, and the decreasing effectiveness of banking regulation," Journal of Finance 36:2, 355-367

Klein, R. W., 2012, "Principles for insurance regulation: an evaluation of current practices and potential reforms," Geneva Papers on Risk and Insurance: Issues and Practice 37:1, 175-199.

Larsen, K., and S. Gilani, 2017. "Regtech is the new black – the growth of regtech demand and investment," Journal of Financial Transformation 45, 22-29

Lenz, R., 2016, "Banking 2025: the bank of the future," Journal of Financial Transformation 44, 111-121

Levitt, T., 1968, "Why business always loses," Harvard Business Review 46:2, 81-89

Lindmark, M., L-F. Andersson, and M. Adams, 2006, "The evolution and development of the Swedish insurance market," Accounting History Review 16:3, 341-370

Lopez, C., 2017, "The asset management industry, systemic risk, and macroprudential policy," Journal of Financial Transformation 45, 121-128

Lopez, L. E., and E. B. Roberts, 2002, "First-mover advantages in regimes of weak appropriability: the case of financial services innovations," Journal of Business Research 55:12, 997-1005

Markham, J. W., 2000, "Banking regulation: its history and future," NC Banking Institute 4:1, 221-286

McIntosh, D., 2016, "The costs of anti-money laundering enforcements to noncompliant banks," Journal of Finance and Bank Management 4:1, 1-14

Mintzberg, H., 1984, "Who should control the corporation?" California Management Review 27:1, 90-115

Moreno, K., 2014, "For the financial sector, regulations are here to stay—time to make the best of them," Forbes, October 9, http://bit.ly/2Hmjn0k

Nelson, R., and S. Winter, 1982, An evolutionary theory of economic change, Harvard University Press

Oliveira, P., and E. von Hippel, 2011, "Users as service innovators: the case of banking services," Research Policy 40:6, 806–818

Oliver, C., and I. Holzinger, 2008, "The effectiveness of strategic political management: a dynamic capabilities framework," Academy of Management Review 33:2, 496-520

Paletta, D., 2017, "Trump administration calls for scaling back post-crisis financial regulations" Washington Post, June 12, http://wapo.st/2GduKrZ

Penrose, E. T., 1959, The theory of the growth of the firm, Oxford University Press

Pilkington, A., and R. Dyerson, 2004, "Incumbency and the disruptive regulator: the case of electric vehicles in California," International Journal of Innovation Management 8:4, 339-354

Richard, P. J., and T. M. Devinney, 2005, "Modular strategies: B2B technology and architectural knowledge," California Management Review, 47:4, 86-113

Rust, S., 2017, "EU considering sustainable investing as fiduciary duty for investors," Investment Pensions Europe, November 13, http://bit.ly/2AHe7lB

Sidhwani, S., 2016, "Seeing the Forest for the Trees – the taming of Big Data," Journal of Financial Transformation 44, 17-19

Son, H., 2015, "JP Morgan CEO Dimon says overlapping regulators assault banks," Bloomberg Business, January 14, https://bloom.bg/2GbwgLc

Tee, R., and A. Gawer, 2009, "Industry architecture as a determinant of successful platform strategies: A case study of the I-Mode Mobile Internet Service," European Management Review 6, 217–232

Teece, D. J., 1986, "Profiting from technological innovation: implications for integration, collaboration, licensing and public policy," Research Policy 15, 285–305

Teece, D. J., 2014, "The foundations of enterprise performance: Dynamic and ordinary capabilities in an (economic) theory of firms," Academy of Management Perspectives 28:4, 328–352

Treleaven, P., and B. Batrinca, 2017, "Algorithmic regulation: automating financial compliance monitoring and regulation using AI and blockchain," Journal of Financial Transformation 45, 14-21

Weber, R. H., 2017, "Regtech as a new legal challenge," Journal of Financial Transformation 46, 10-17

Wessel, D., 2012, "Bank balance: regulation and innovation," Wall Street Journal, April 4

Wiseman, R. M., and C. Catanach, 1997, "A longitudinal disaggregation of operational risk under changing regulations: evidence from the savings and loan industry," Academy of Management Journal 40:4, 799–830

Wolf, A., and W. Huber, 2017, "Banking regulation 2017: Austria," Global Legal Insights, http://bit.ly/2FDGrdM

Worstall, T., 2015, "The next great problem: unregulated shadow banking now worth $80 trillion globally," Forbes, November 13, http://bit.ly/2Hmbzf6

ABOUT CAPCO

Capco is a global technology and management consultancy dedicated to the financial services industry. Our professionals combine innovative thinking with unrivalled industry knowledge to offer our clients consulting expertise, complex technology and package integration, transformation delivery, and managed services, to move their organizations forward. Through our collaborative and efficient approach, we help our clients successfully innovate, increase revenue, manage risk and regulatory change, reduce costs, and enhance controls. We specialize primarily in banking, capital markets, wealth and investment management, and finance, risk & compliance. We also have an energy consulting practice. We serve our clients from offices in leading financial centers across the Americas, Europe, and Asia Pacific.

To learn more, visit our web site at **www.capco.com, or follow us on Twitter, Facebook, YouTube, LinkedIn and Xing.**

WORLDWIDE OFFICES

Bangalore
Bangkok
Bratislava
Brussels
Charlotte
Chicago
Dallas
Dusseldorf
Edinburgh
Frankfurt
Geneva
Hong Kong
Houston
Kuala Lumpur
London
New York
Orlando
Paris
Pune
São Paulo
Singapore
Stockholm
Toronto
Vienna
Warsaw
Washington, DC
Zurich

CAPCO.COM

CAPCO